NASA CONTRACTOR REPORT

NASA CR-1888

NASA CR-

DEVELOPMENT OF TECHNIQUES FOR MEASURING PILOT WORKLOAD

by D. A. Spyker, S. P. Stackhouse, A. S. Khalafalla, and R. C. McLane

Prepared by
HONEYWELL, INC.
Roseville, Minn. 55113
for Ames Research Center

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • NOVEMBER 1971

1. Report No. NASA CR-1888	2. Government Accession No.	3. Reci ... No.
4. Title and Subtitle Development of Techniques for Measuring Pilot Workload		5. Report Date November 1971
		6. Performing Organization Code
7. Author(s) D.A. Spyker, S.P. Stackhouse, A.S. Khalafalla and R.C. McLane		8. Performing Organization Report No.
		10. Work Unit No.
9. Performing Organization Name and Address Honeywell Inc. Systems and Research Center 2345 Walnut Street Roseville, Minnesota		11. Contract or Grant No. NAS 2-5443
		13. Type of Report and Period Covered Contractor Report
12. Sponsoring Agency Name and Address National Aeronautics & Space Administration Washington, D.C. 20546		14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

The goal of this study was to provide an objective method of assessing information workload based on physiological measurements. Information workload or reserve capacity, was measured using a visual discrimination secondary task and subjective rating of task difficulty. The primary task was two axis (pitch and roll) tracking, and the independent variables in this study were aircraft pitch dynamics [K/S, K/S^2, and $K(S + 1)\ 16/S\ (S^2 + 8S + 16)$] and wind gust disturbances (white noise with cut-offs at 1.5, 2.5, and 4.0 rad/sec.). The study was structured to provide: 1) a sensitive, nonloading mesure of reserve capacity, and 2) an unencumbering reliable measurement of the psychophysiological state. From these, a measured workload index (MWI) and physiological workload index (PWI) were extracted. An important measure of the success of this study was the degree to which the MWI and PWI agreed across the 243 randomly-presented, four-minute trials (9 subjects X 9 tasks X 3 replications).

The electrophysiological data collected included vectorcardiogram, respiration, electromyogram, skin impedance, and electroencephalogram. Special computer programs were created for the analysis of each physiological variable. The digital data base then consisted of 82 physiological features (e.g., heart rate, respiration rate, etc.) for each of the 243 trials.

A prediction of workload based on physiological observations was formulated as a simultaneous least-squares prediction problem. A "best" subset of 10 features was chosen to predict the three measures of reserve capacity. The cannonical correlation coefficient was .754 with a chi squared value of 91.3 which allows rejection of the null hypothesis with p .995.

17. Key Words (Suggested by Author(s)) Information Workload, Pilot Workload, Reserve Capacity	18. Distribution Statement UNCLASSIFIED-UNLIMITED

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
UNCLASSIFIED	UNCLASSIFIED	114	3.00

*For sale by the National Technical Information Service, Springfield, Virginia 22151

FOREWORD

This report describes the work accomplished at Honeywell Inc. on the pilot workload measurement research program under Contract NAS2-5443 for the period 1 June 1969 through 15 June 1970.

The study was conducted for the National Aeronautics and Space Administration, Ames Research Center, under direction of Mr. Miles R. Murphy, Technical Monitor.

We wish to acknowledge Darryl A. Erlien, Clinton L. Jolliffe, and Joan Bleedorn for their assistance in executing this study.

CONTENTS

CONTENTS - Concluded

ILLUSTRATIONS

TABLES

DEVELOPMENT OF TECHNIQUES FOR MEASURING PILOT WORKLOAD

By D. A. Spyker, S. P. Stackhouse,
A. S. Khalafalla and R. C. McLane

SUMMARY

Virtually every critical aspect of civilian and military aircraft operation involves a human display/control system interaction. An evaluation of these display/control systems which is based exclusively on operator performance is addressing only a fraction of the problem. That is, a pilot with one display configuration may work twice as hard (twice the workload) as he does with another, yet achieve equal performance for both. It would be of particular value to have an efficient technique which provides a quantitative, objective measure of operator workload. Such a measure, when used in addition to performance data, would permit a significant improvement in the evaluation of the effects of environmental stress and time line selection of tasks, as well as alternative control/display configurations.

The development and evaluation of such a technique for workload measurement was the objective of this effort.

The goal of this study was to provide an objective, quantitative method of measuring pilot workload based on electrophysiological measurements. The main subproblems in this study were:

1) Validation of a sensitive, nonloading secondary task for evaluating the subject's reserve capacity

2) Collection of physiological and performance data over a range (easy to hard) of visual motor tracking tasks

3) Extraction of any potentially meaningful features from the analog physiological data

4) Normalization of the features

5) Selection of the "best" subset of these features

6) Simultaneous computation of the workload index and the best linear predictor from the subset of features

7) Validation of this predictor.

The operator's psychophysiological state is explicitly influenced by his physical and psychological condition and environment, which were held constant as far as possible during the study. The psychophysiological state determined the total capacity available for the performance of the visual/motor

task. Based on the subject's performance some fraction of the total capacity was expended. The difference, or reserve capacity, was measured using the visual discrimination secondary task and subjective rating of task difficulty. The primary task was two-axis (pitch and roll) tracking, and the independent variables in this study were aircraft pitch dynamics $[K/S,\ K/S^2$, and $K(S+1)16/S(S^2+8S+16)]$ and wind gust disturbances (white noise with cut-offs at 1.5, 2.5, and 4.0 rad/s).

The entire study was structured to provide: 1) a sensitive, nonloading measure of reserve capacity, and 2) an unencumbering, reliable measurement of the psychophysiological state. From these a measured workload index (MWI) and a physiological workoad index (PWI) were extracted. An important measure of the success of this study was the degree to which the MWI and PWI agreed across the randomly-presented 243 four-minute trials (9 subjects x 9 tasks x 3 replications).

This study provided three direct measures of reserve capacity:

1) Miss Rate - Percent of error in responding to the secondary task

2) Response Time - Average time from secondary task stimulus onset to response

3) Subjective Rating - Pilot's evaluation of task difficulty.

All three of these were found sensitive to workload.

The approach to finding electrophysiological parameters which are sensitive to workload consisted of:

1) Using a carefully designed multichannel physiological monitoring system

2) Using an automatic, digital computer feature extraction system

3) Using a pattern recognition system approach to the selection of the "best" subset of features.

The electrophysiological data which were collected included:

- Vectorcardiogram
- Respiration
- Electromyogram
- Skin impedance
- Electroencephalogram (visually evoked cortical response).

The analog data base was converted to a digital data base by sampling, extracting features, and writing a digital magnetic tape record for each session. A special program was created for the analysis of each physiological

variable. The digital data base then consisted of 82 physiological features (e. g., heart rate, respiration rate, etc.) for each of the 243 trials. A critical step in the study was the normalization of the data. Each feature was represented as its percent change from the experimental session average (since each of the three sessions included all nine tasks), and then an across-replication average was taken. The results were all referred to this normalized data base.

Based on either performance or measured workload, the three easiest and three hardest tasks represented two fairly distinct classes. It thus seemed appropriate to use the a priori class membership in a two-class pattern recognition study of the physiological data base. The pattern recognition program performs cluster seeking, feature selection, discriminant design, and classification based on a least-squares criterion. For eight features the separation was 96. 3 percent. That is, based on the physiological features, subjects could be correctly classified as performing easy versus hard tasks 96. 3 percent of the time.

Three sensitive measures of workload were described, and it was shown that certain of the physiological features permitted good discrimination between easy and hard tasks. The final objective was to formulate an accurate, reliable prediction of workload based on electrophysiological observations. Since it was not obvious which of the several measures were best, the question was formulated as a simultaneous least-squares prediction problem.

- Given: The m measures of workload ($y_1, \ldots y_m$) and the n physiological features ($x_1, \ldots x_n$).

- Find: The m+n coefficients such that the measured workload index

  $$MWI = b_1 y_1 + b_2 y_2 + \ldots + b_m y_m$$

  is best predicted by the physiological workload index

  $$PWI = a_1 x_1 + \ldots + a_n x_n$$

  i. e.,

  $$\sum_{i=1}^{N} (PWI - MWI)^2$$

  is minimized over the (N=81) trials.

Through a combination of classification ordering and multiple correlation ranking a "best" subset of 10 features was chosen to predict miss rate and response time, the secondary task measures of reserve capacity. The Cannonical Correlation coefficient was .646, and solutions for the coefficients were found.

The chi squared value, 57.5 with 20 degrees of freedom, allowed rejection of the null hypothesis with $p > .995$.

Application of these weights to a separate set of validation data resulted in a correlation coefficient between MWI and PWI of .569. To estimate the significance of this result, MWI and PWI may be considered as simply $N = 20$ pairs of points. The null hypothesis was rejected with $p > .99$.

Although subjective rating was not used in the validation study, it was a sensitive measure of workload. If it was also included in the MWI, the value of the correlation coefficient increased to .754 with a chi squared value of 91.3, with $p > .995$.

The salient features of this study which represent new or substantially improved techniques include:

1) A simple, sensitive, nonloading secondary task

2) A subjective rating which agrees with other secondary task measures, but with less intersubject variance

3) A multichannel physiological monitoring and recording system for respiration, vectorcardiogram, electromyogram, electroencephalogram, skin impedance, and subject performance

4) Automatic feature extraction software which transforms the analog data base into meaningful features

5) Very good separation results using a pattern recognition system, assuming the data to represent a two-class problem

6) Use of simultaneous least-squares prediction to arrive at a statistically significant, validated workload index and the physiological features which best predict it.

INTRODUCTION

The selection of alternative aircraft subsystem configurations is a common and important design problem. From a human engineering standpoint, it is desirable to know which one of several candidate configurations permits the best use of human performance capabilities. An evaluation procedure which relies exclusively on performance measures is inadequate. That is, a pilot with one configuration may work twice as hard as he does with another, yet achieve equal performance for both. Thus, one can conclude that the pilot's capabilities were unequally taxed and that this inequality was not detected by a performance measurement. It follows that a proper experimental design should include some method of assessing the amount of the pilot's capability which was used in obtaining a given level of performance.

Reserve Capacity

To accomplish this measurement, it is common (Brown, 1964; Knowles, 1963; Hilgendorf, 1965) to postulate a construct known as reserve capacity. While this construct can be applied in a physical workload sense, its value for our purposes is in the context of information workload. To measure reserve information processing capacity, a second task is imposed on the pilot or other subject. The extent to which the subject can satisfactorily perform the secondary task, while still performing adequately on the primary task, is taken as a measure of his reserve capacity. (Measures of reserve capacity are discussed more thoroughly in Appendix A.)

An adequate secondary task for our experiments has two attributes which must be satisfied before it can be considered as a measure of reserve capacity. First, it must be sensitive to primary task difficulty, i.e., as the primary task becomes more difficult (error increases), the number of errors on the secondary task must increase. Second, the secondary task must not load the primary task. That is, performance on the secondary task must not cause primary task errors. This second requirement for the secondary task introduces the difficult question of motivation.

There are two general approaches for solving the motivation problem. Either a carefully adjusted schedule of appropriate reinforcement can be employed, or subjects can be selected on the basis of their ability to satisfy the requirements imposed by the secondary task. The former approach requires a delicate adjustment of type and frequency of payoff for the entire population of subjects, while the latter approach only requires an adequate population from which well-motivated subjects can be selected.

Once a measure of reserve capacity has been established, experiments can be designed to determine the extent of the subject's capabilities which are used in performing tasks of varying degrees of difficulty. Thus, the workload measures which are obtained can be accurate, reliable, and internally consistent within the frame of reference provided by the particular experimental situation. The next question which arises is how well the findings obtained in the laboratory generalize to a real-world operational situation. If the real-world environment is at all complex, it is immediately apparent that the laboratory experiment cannot be directly validated under actual operating conditions. For example, a tracking task which simulates manual control of an aircraft, plus a secondary task to measure reserve capacity, cannot be implemented in the cockpit of an aircraft which will be flown by pilots who are evaluating alternative subsystem configurations. Such a direct validation of the laboratory findings cannot be accomplished. Instead, measures of variables must be taken which correlate with the variables measured in the laboratory performance experiments. In addition, the new set of variables must not interfere with the pilot's primary responsibility, i.e., flying the airplane.

Psychophysiological Variables

In past experiments, as well as in the present study, psychophysiological variables have been measured and correlated with some aspect of a subject's performance. (Appendix B discusses some psychophysiological measures of performance.) The general finding in this area has been that if the information workload demands placed on the subject are very different in degree of difficulty, an indication of this difference can be obtained by an analysis of psychophysiological variables. However, if the tasks are not widely different in degree of difficulty, significant physiological differences typically are not found (e.g., Jex and Allen, 1970). The lack of positive findings is usually attributed to the lack of sensitivity or inappropriateness of the physiological measures which were used. Another possible cause of this problem may at times be traced to shortcomings in analysis of the physiological data. It is reasonable to suppose that there is more information in the analog physiological data than can be obtained from an analysis which is limited to measurements of amplitudes and intervals. It is likely that a more thorough analysis of the data might yield results which correlate with performance measures.

Subjective Evaluation

Subjective evaluations of performance have been widely used by a number of investigators. The topic is complex and the merits and difficulties of subjective performance or workload evaluations have been discussed at length (e.g., McDonnell, 1968 and 1969). While subjective techniques are generally considered to lack reliability and precision, they are often the best method for measuring workload in the operational environment. This may be due in part to the difficulty of applying laboratory instrumental techniques in the field.

Description of This Study

In this study we have evolved a workload index based on the pilot's physiological response to a simulated tracking task. Important steps in this approach include:

1) Validation of a sensitive, nonloading secondary task

2) Collection of physiological and performance data over a range (easy to hard) of visual/motor tracking tasks

3) Extraction of any potentially meaningful features from the analog physiological data

4) Normalization of the features

5) Selection of the "best" subset of these features

6) Simultaneous computation of the workload index and the best linear predictor from the subset of features

7) Validation of this predictor.

The structure of this study is summarized in Figure 1. The operator's psychophysiological state is explicitly influenced by physical condition, psychological condition, and stress, as well as implicitly effected by expended capacity (E. C.) and performance.

The psychophysiological state determines total capacity (T. C.), and that fraction which is unused is called reserve capacity (R. C.); i. e. :

$$T.C. = E.C. + R.C.$$

From the subjective rating and secondary task performance a measured workload index (MWI) was extracted, and from the electrophysiological data a physiological workload index (PWI) was extracted. A final measure of the success of the study was the agreement between MWI and PWI across subjects and across trials.

PRIMARY TRACKING TASK

The objective of this study was the development of a technique for assessing pilot workload (or reserve capacity) based on psychophysiological measurements. Since the measurement of reserve capacity on an absolute scale is difficult at best, it was necessary to examine the change in reserve capacity as a function of primary task difficulty. Thus, the tracking task represents the independent variable in this study, and our main concern was that it provide the pilot with a broad range of difficulties (workloads) and corresponding changes in reserve capacity.

The primary task chosen for this study was two-axis tracking (pitch and roll) with a CRT compensatory display and displacement side stick.

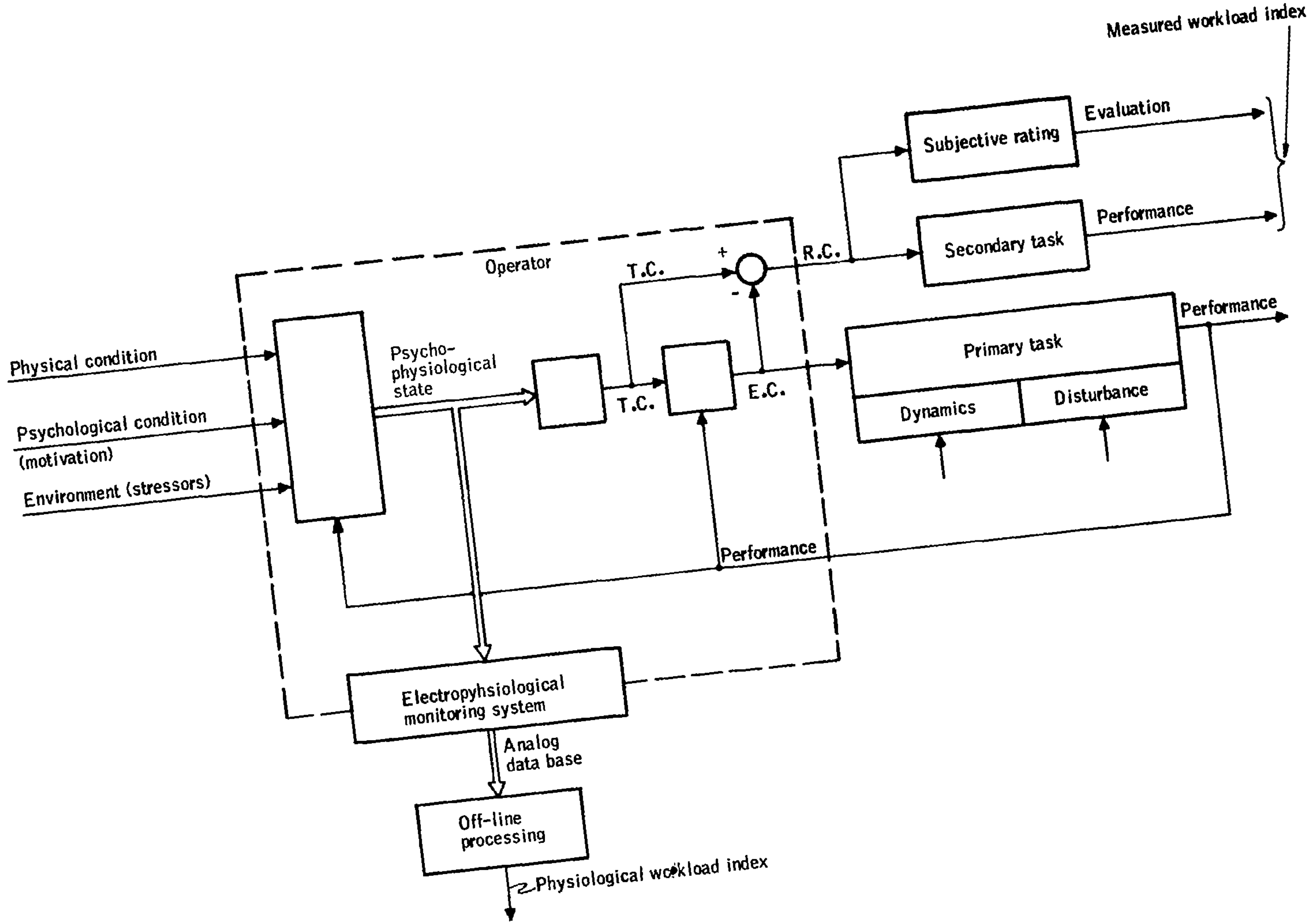
Dynamics

The stationary dynamics for the roll axis were K/S throughout the study, while three pitch dynamics were used: K/S, $K(S+1)\ 16/S\ (S^2+8S+16)$, and K/S^2.

Display

The display was generated on a standard 8-cm x 10-cm CRT* set in a flat black plywood panel which also contained the strobe light and discrimination lights (Figure 2). The remainder of the pilot booth (4 ft x 6 ft) was formed of black curtain material.

*Tektronix RM 561A.



1 Simplified Experimental Structure

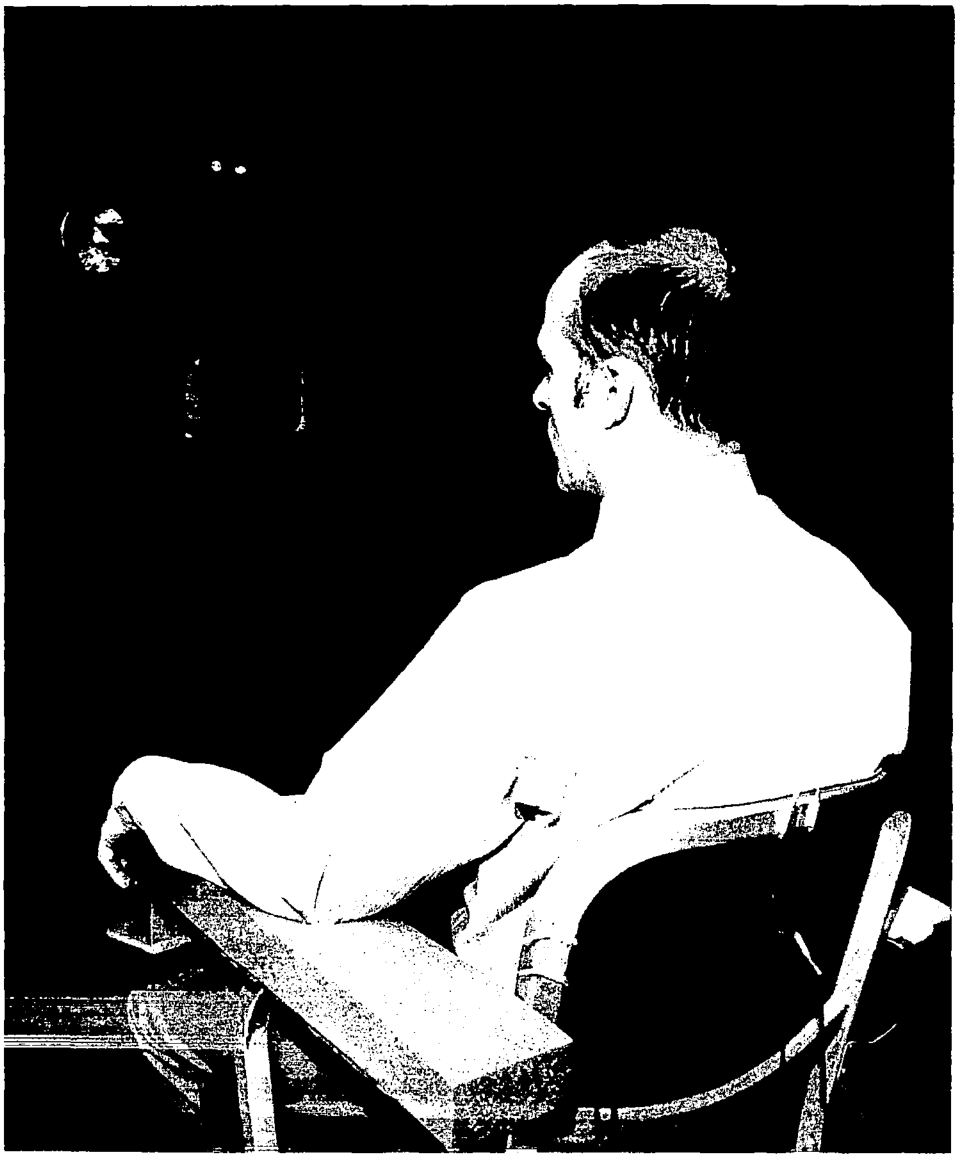


Figure 2. Subject's View of Display Used in Experiment

The display scale, 10°/cm, at a viewing distance of 76 cm, was compressed from the 0.75°/cm given in contact flight to the earth horizon but was still within range for proper use of the small-angle approximations used in generating the display. The display was "inside-out", i.e., an artificial horizon which moves up when the aircraft pitches down.

Since pilots were instructed to respond to the secondary task lights only if they felt they could do so without degrading their tracking performance, it was necessary to provide a no-penalty display region. Reticles were provided on the CRT face at ±3.75° of pitch and ±2.8° of roll error (Figure 3). Tracking error in excess of these limits was squared and accumulated (integrated) for each axis.

Pitch and roll were limited to ±30° to prevent loss of display cursor and consequent data loss.

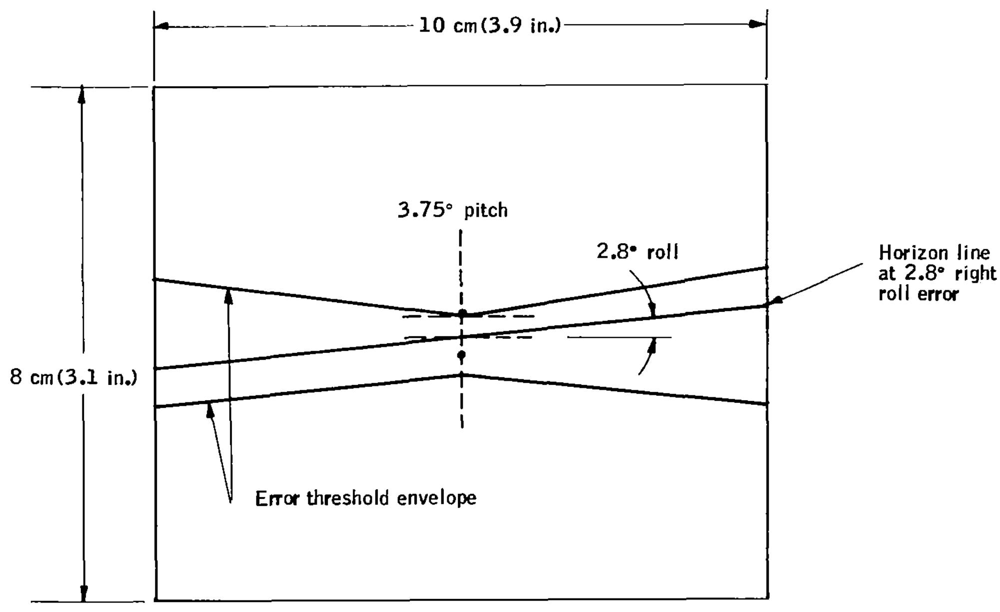


Figure 3. Compensatory Two-Axis Display

Control

The hand control was a right-hand side stick (all pilots were right handed) with ±45° of roll freedom and ±25° of pitch freedom. It is a 400-Hz variable transformer displacement stick without spring centering or detent. Maximum stick displacement for the K/S dynamics corresponds to pitch and roll rates of 30°/s, and the stick is essentially linear over its operating range.

Forcing Function

The input forcing function was gaussian white noise with second-order filter cutoffs at 1.5, 2.5, and 4.0 rad/s. Independent generators* and filters were used, but both pitch and roll received noise with the same cutoff. The noise amplitude was the equivalent of 7.5° rms.

The tracking simulation is summarized in Figure 4.

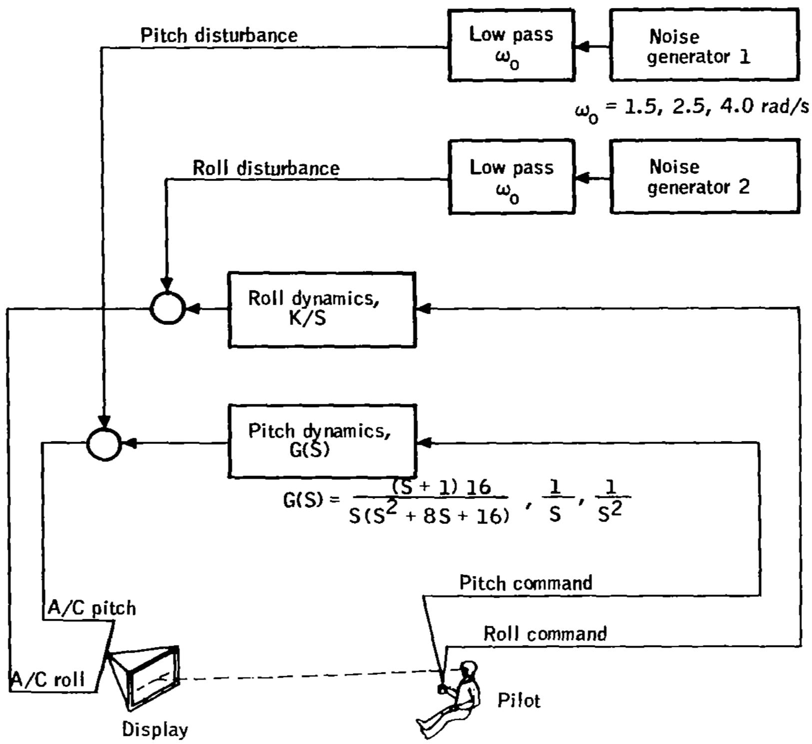


Figure 4. Summary of Tracking Simulation

Experimental Design

Main experiment. - The experimental design was a 3 x 3 factorial design in the tracking task (independent variable), using the matrix shown in Table I.

The numbers in the grid represent a task ranking by tracking performance, hereafter referred to as task numbers.

---

*Pace 44.200 low-frequency gaussian noise generator.

The nine subjects for the study were liscensed pilots and each was practiced until his tracking score: a) reached an acceptable level, and b) assymptoted.

TABLE I. - FACTORIAL DESIGN (MAIN EXPERIMENT)

Noise cutoff	Pitch dynamics		
	$\frac{K(S+1)16}{S(S^2+8S+16)}$	$\frac{K}{S}$	$\frac{K}{S^2}$
1.5	1	2	7
2.5	3	4	8
4.0	5	6	9

Several subjects failed to meet these criteria and were rejected. Each subject performed three replications of each of the nine tasks for the main experiment. These 27 runs were divided into three sessions of nine runs each, with prebaseline and postbaseline data taken (Figure 5). The sessions were conducted, as nearly as possible, on three consecutive days and each session included all nine tasks in a random order (see Appendix C). This feature is important since it permitted normalization of data based on session averages.

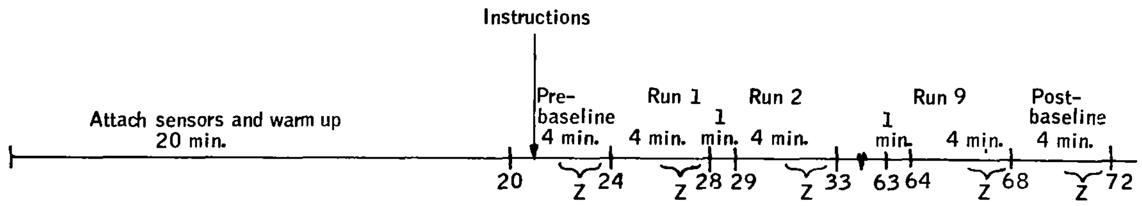


Figure 5. Time Line for Main Experiment (Z Denotes Skin Impedance Measurements)

Validation study. - The validation study was undertaken for two reasons:

1) To provide an independent data base to check the workload index
2) To study the effects of a step change in pitch dynamics.

For this study, only the center noise cutoff frequency (2.5 rad/s) was used. The subject was presented with either K/S or K/S^2 for three minutes and then the pitch dynamics were switched to the other for three minutes (Figure 6). The random presentation design is included in Appendix C.

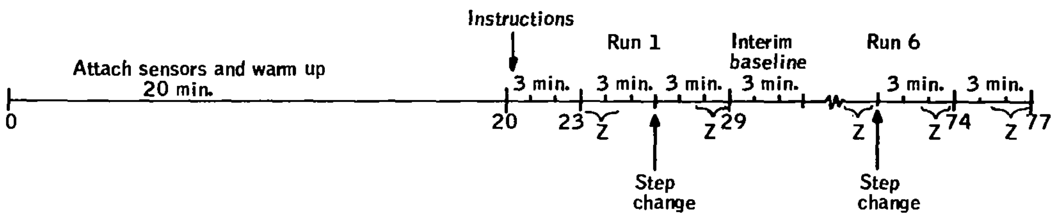


Figure 6. Time Line for Validation Study (Z Denotes Skin Impedance Measurements)

Tracking Performance

Tracking error (TE) is defined as

$$TE = \frac{1}{T}\left[\left(\int_0^T E_P(t)^2\, dt\right)^{1/2} + \left(\int_0^T E_R(t)^2 dt\right)^{1/2}\right]$$

where

T = duration of the run = 4 minutes

E_P = error in pitch beyond ±3.75°

E_R = error in roll beyond ±3.8°

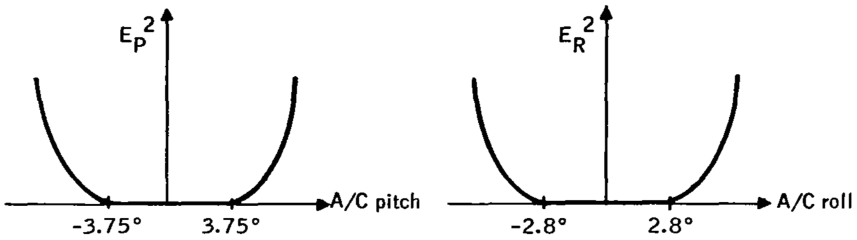


Figure 7 presents the mean and standard deviations* of the average tracking error by task number. Each point is the average of 27 points (three replications for each of nine subjects). From this figure it is apparent that there is a large variance in the across-subject performance averages. It also seems to illustrate the limitations of performance as an indicator of workload; i. e., although tasks 1 through 5 are certainly increasing in difficulty, they do not show the kind of performance degradation evident in tasks 6 through 9. This is in clear contrast to the secondary task measures described in the following section.

For the validation study, separate primary and secondary task measures were made before and after the step change which occurred in the center of the six-minute run. Task numbers for the validation study were assigned as follows:

Task Number	Dynamics	Presented
11	K/S	Before step change
12	K/S	After step change
13	K/S^2	Before step change
14	K/S^2	After step change

The mean and standard deviations for these tasks are presented in Figure 8, and collected in Appendix D. This figure illustrates the expected difference between the K/S and K/S^2 dynamics and the relatively subtle increase in error which results from presenting the dynamics as a step change.

Summary

More or less arbitrarily, a compensatory two-axis tracking task with three pitch dynamics and three disturbance cutoff frequencies was selected to provide the independent variable for this study. Results indicate that these

*We use $\sigma = \sqrt{\frac{1}{N} \Sigma X^2 - \bar{X}^2}$.

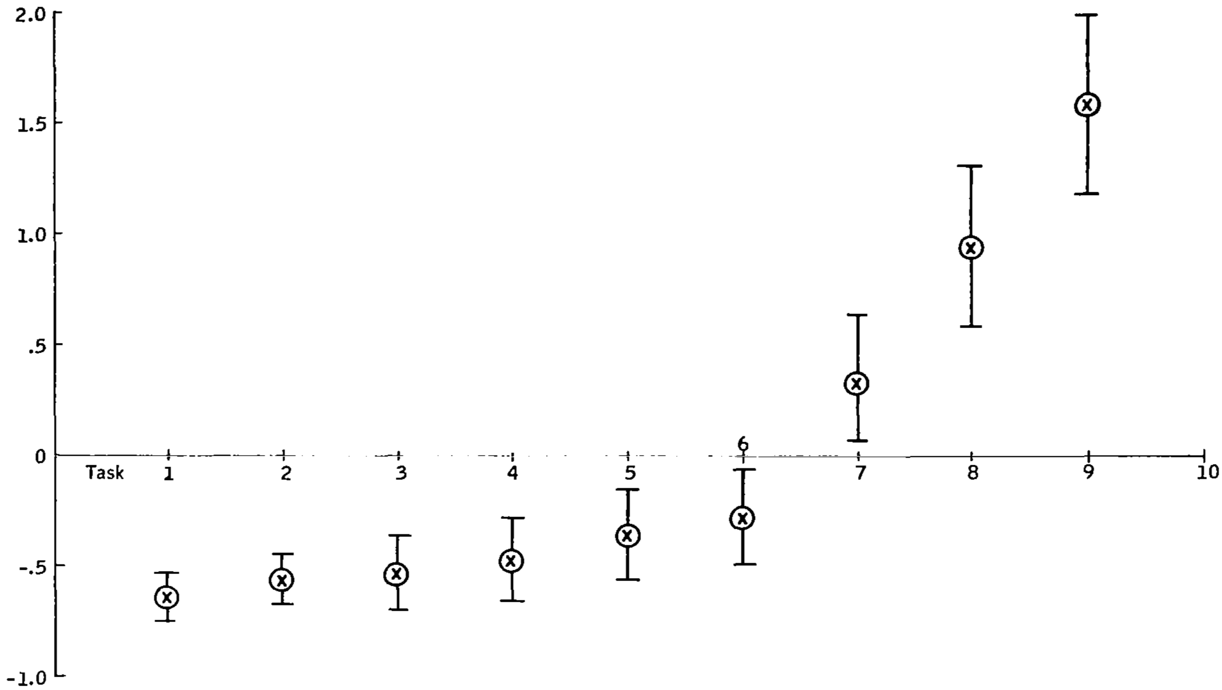


Figure 7. Tracking Error by Task - Main Experiment

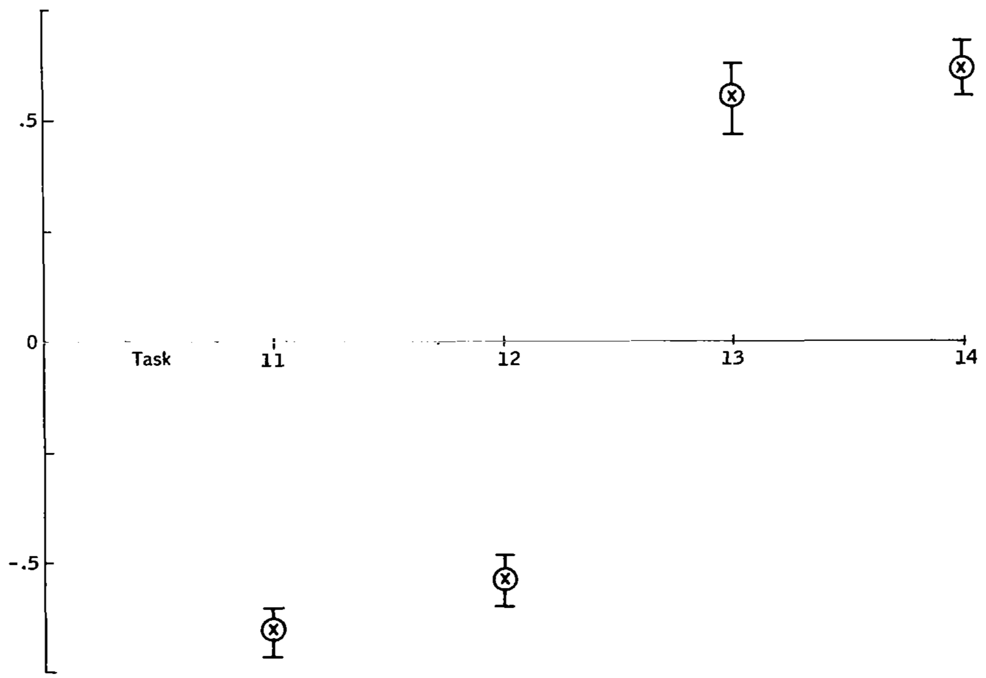


Figure 8. Tracking Error by Task - Validation Study

nine tasks span both the regions where workload change is more pronounced than performance and where performance changes more than workload (approaching 100 percent workload).

SECONDARY (DISCRIMINATION) TASK

Since the goal was to develop an objective, quantitative predictor of reserve capacity based on psychophysiological measures, it was necessary to find some independent reserve capacity measurement. To provide this measure, a subsidiary visual/motor discrimination task was selected (the rationale of this approach is detailed in Appendix A). The two principal criteria for selection of the task for this study were:

1) Minimum loading of primary task by the secondary task, where loading is defined as degradation of tracking performance with addition of the secondary task

2) Maximum sensitivity of secondary task performance to primary task workload.

The stimulus was provided by a pair of 0. 5-inch-diameter lights with 1. 5-inch horizontal separation. The subject responds (left or right) via a thumb-operated rocker switch* mounted in the top of the control stick (see Figure 2).

Demand Task

During the preliminary study, both self-paced (demand) and random presentations were evaluated. In the former approach, the secondary task is intended to use the subject's entire reserve capacity. For the demand presentation, the light remained on until the correct response. After a 300-ms delay, the right or left lamp was lit with equal probability.

For this configuration, the lamps were placed directly over the display, out of foveal vision, and out of peripheral vision. This approach was abandoned due to the loading effects (degradation of tracking performance) which ranged from 20 percent to 70 percent.

Random Presentation

For random presentations of the lights, a recorded stimulus with a certain mean (m_r) and standard deviation (σ_r) was used. At the occurrence of a stimulus, the right or left lamp (with equal probability) was lit for T_s seconds, and a correct response was counted if it occurred within T_r seconds. If the subject latency is defined to be T_i for the i^{th} stimulus, thenthe response was

---

*A special short-throw switch was intalled which has a 2-mm excursion, "on" pressure of 50 grams, and opens at 20 grams.

correct if $T_i < T_r$. In addition, cumulative latency (ΣT_i) was computed for each run.

The following parameters were adopted for the secondary task:

- Rate (m_r, σ_r) — 3 ± 0.5 sec
- Duration $T_s = T_r$ — 800 ms
- Location — 12 in. up from display
- Luminance — Relatively dim, 14.5 ft-L

The stimulus for the secondary task was prerecorded on a single-channel Wollensak tape recorder, and the right/left light was determined by the state of a flip-flop which was toggled at 4 kHz when the lights were off. The logic (asynchronous) and stimulus and response counters are included in a general-purpose, Honeywell-built logic rack. Right and left stimuli, as well as correct responses right and left, were separately accumulated on mechanical counters, while cumulative latency was recorded to the nearest 0.01 second by a Beckman counter. The subject's first response was the only one considered, since the logic inhibits the opposite response for 800 ms.

The preliminary studies indicated that this configuration provided the required sensitivity to workload changes with minimal loading (5 percent).

It should be pointed out that this discrimination task was subsidiary by pilot decision. To ensure that these instructions were clear, a set of tape-recorded instructions were played during the prebaseline portion of each session (see Appendix C).

Workload Measures

The two measures of secondary task performance which were subsequently used in predictor development are miss rate and response time:

$$\text{Miss rate, MR} = \frac{\text{Number stimuli missed}}{\text{total number stimuli}} \times 100 \text{ percent}$$

$$\text{Response time, RT} = \frac{\text{Cumulative latency } (\Sigma T_i)}{\text{Total number stimuli}}$$

During the preliminary study, approximate values of MR = 5 percent and RT = Σ490 ms were established for these parameters without the primary task, but with the subjects looking at the tracking display.

Pilot Secondary Task Performance

Complete tabulation of pilot performance on the secondary task is included in Appendix D, but the results are summarized in the following figures. Figure 9 shows the across-subject means and standard deviations by tasks. This figure shows the increase in workload of tasks 1 through 5 even though the tracking performance (Figure 7) was nearly constant. In contradistinction, these measures suggest a comparable workload for tasks 7 through 9, whereas tracking error is steadily increasing over the same tasks.

From this we conclude that the first six tasks represent increasing workloads as well as increasing difficulty but that the three hardest tasks represent maximum workloads for this experimental condition, and thus tracking errors increase with the difficulty. The fact that this assymptote occurs at 25 percent miss rate suggests that the secondary task was still easily performed at a high workload level and that perhaps the secondary task should have been more difficult.

Figure 10 shows the secondary task performance during the validation study. Although the miss rate does not seem sensitive to the order of presentation (in task 11, K/S is presented first, and in task 12 it occurs after the step change from K/S^2), the response time does appear sensitive to order.

Although the across-subject variance seems large, this is largely attributable to subject differences. For example, RT for task 13 is 600 ± 93, while the individual pilot's RT ranged from 554 ± 4 to 670 ± 30.

Correlations

For each of the nine tasks there are three replications for each subject. If these replications are averaged, then there are 81 data sets in the main experiment. The correlation coefficient

$$R = \frac{\overline{xy} - \bar{x}\bar{y}}{\sigma_x \sigma_y}$$

where

$$\bar{x} \triangleq \frac{1}{N} \sum_{i=1}^{N} x_i$$

provides a simple scalar measure or the linear relationship between two variables.

The correlation coefficient matrix shown in Table II summarizes the relations between tracking secondary and subjective rating scores (described in the following section).

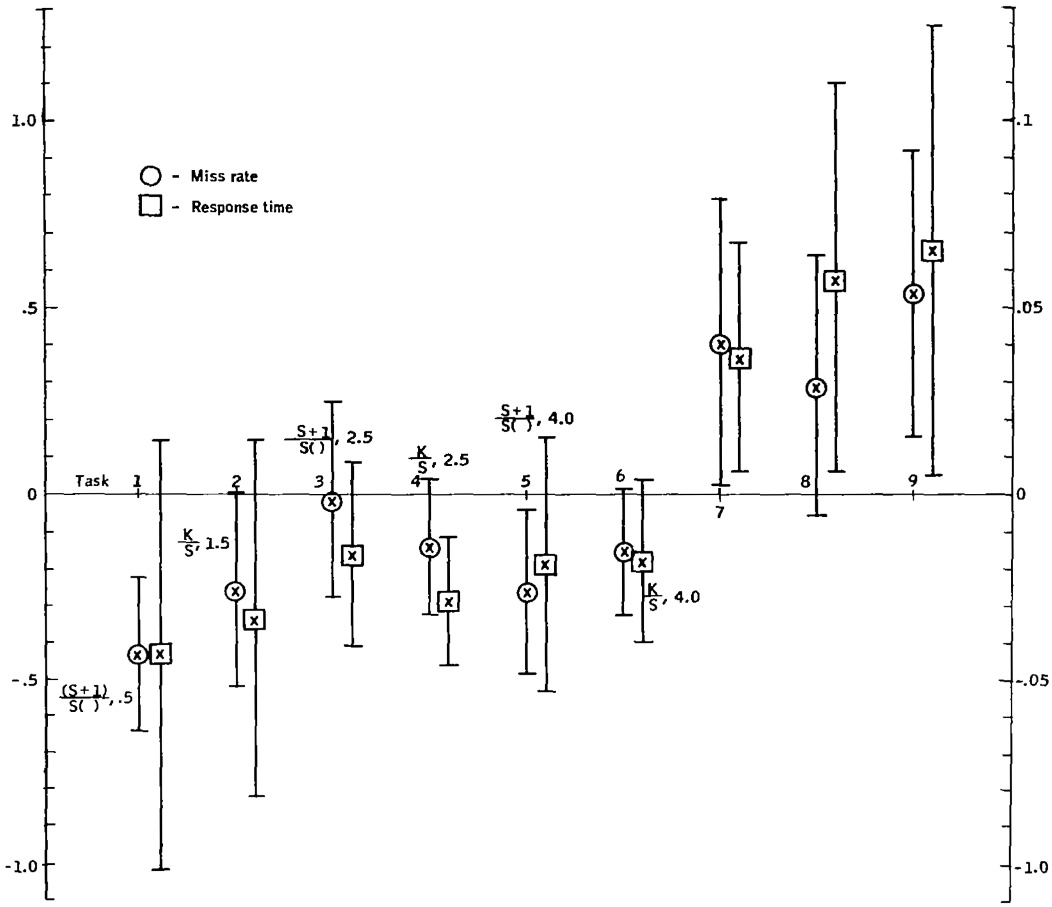


Figure 9. Miss Rate and Response Time - Main Experiment

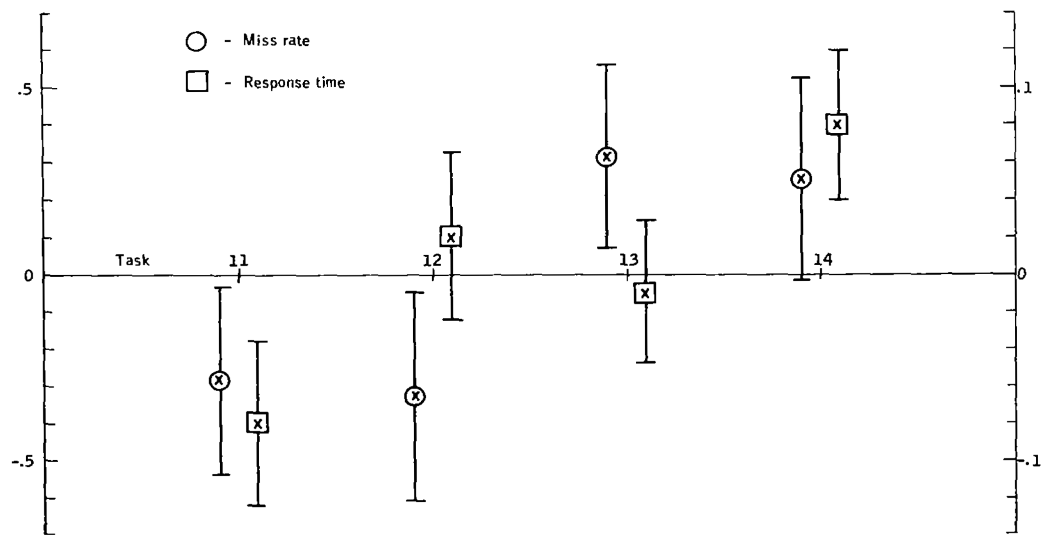


Figure 10. Miss Rate and Response Time - Validation Study

TABLE II. - CORRELATION COEFFICIENTS FOR UNNORMALIZED SCORES

Feature	Tracking error	Miss rate	Response time	Subjective rating
Tracking error	1.00	.298	.215	.600
Miss rate	.298	1.00	.860	.590
Response time	.215	.860	1.00	.472
Subjective rating	.600	.590	.472	1.00

The 5-percent and 1-percent significance levels for N = 81 are $R_{5\%}$= .217 and $R_{1\%}$ = .283; i.e., the null hypothesis (R = 0) is rejected at the .01 level for R $\geq$ .283. It is reassuring to observe that miss rate and response time are highly correlated with each other and with subjective rating, although subjective rating is more highly correlated with tracking error. The low correlations of TE with MR and RT supports the premise that primary task performance is a poor predictor of workload.

Conclusions

The randomly presented parallel discrimination task (also visual/motor as is the primary task) provides a nonloading, reasonably sensitive measure of the primary tracking task workload in this experiment. The secondary task performance suggests that the workload increases approximately linearly in tasks 1 through 6 and that tasks 7 through 9 (acceleration control) represent 100 percent workload in this experiment.

SUBJECTIVE EVALUATION

Objective

A measurement of each subject's evaluation of the primary and secondary task difficulty was included to provide an independent measure of workload.

Selection of Questionnaire

The questionnaire which was used (Appendix C) had four multiple-choice questions related to the difficulty of the primary task and two questions on the difficulty of the secondary task. Points were assigned for each answer on a zero-to-10 scale, with 10 indicating the greatest difficulty or highest workload.

Subjective Evaluation of Task Difficulty

The subject of rating scales for handling qualities for both real and simulated aircraft performance has received considerable attention in recent years. McDonnel (1968 and 1969) critically reviewed this topic. One of his conclusions is interesting and pertinent to this program. He observes that contemporary scales are the result of a lengthy trial and error development. This process has led to the use of scales which are difficult to improve. While this finding does not justify our rating procedure, it does tend to justify our use of portions of the Cooper-Harper scale, as well as the questions which we devised to correspond to the Cooper-Harper format. The most encouraging finding is that the scale we used corresponded to both primary and secondary task difficulty in our simulation. That is, on the tasks judged as most difficult by the subjects, the most errors were made on both primary and secondary tasks. This finding, we believe, justifies the inclusion of our rating scale in the main experiment.

Our scale was limited to six questions due to the time restrictions in the experimental design. A lengthy questionnaire would have unduly prolonged the experimental session.

Results

The subjective evaluation forms were used only during the main experiment, since there were no rest periods between tasks during the validation experiment. The across-subject means and standard deviations for all six applications of the questionnaire and for all main experiment subjects are shown in Figure 11. For these data, maximum difficulty would receive a score of 60. It is clear that as the task became more difficult, the subjective evaluation of difficulty also increased.

PHYSIOLOGICAL MONITORING

One of the salient features of this study is that it represents simultaneous monitoring and analysis of multiple physiological variables, including electrocariogram, electromyogram, respiration, electroencephalogram, and skin impedance. It is obvious that a prerequisite to a successful study is a consistently good physiological monitoring system.

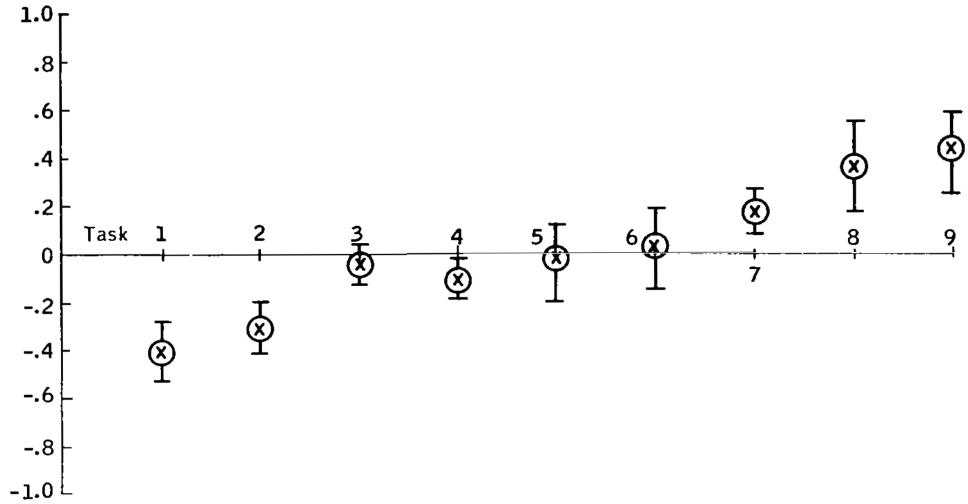


Figure 11. Subjective Rating Task Averages - Main Experiment

In this section, the psychophysiological variables selected for this study are described, as is the creation of the analog data base (FM magnetic tape).

The physiological observations which we initially considered were:

1) Electromyogram (EMG)
2) Respiration
3) Vectorcardiogram (VCG)
4) Skin impedance
5) Electroencephalogram
6) Eye movement

After preliminary investigations only eye movement was eliminated and this was due to a combination of lack of promising results, difficulty of extraction, and subject discomfort of the electrodes.

Electromyogram

The surface electromyogram is the potential generated by the contraction of muscle fibers. This potential typically ranges from 0. 1 to 1 mV and can be reasonably represented as amplitude-modulated noise (Kreifeldt, 1969).

Our interest in the EMG was as an indicator of tension in a noninvolved muscle, and we tried several locations on the neck and the off-side (left) arm. The consistency of results and ease of eliminating ECG artifact favored differential amplification of signals from the belly of the brachial radialis and an electrode opposing it over the ulna. This provided for a signal which was almost exclusively due to brachial radialis contraction, since the electrode over the bone approximated a passive reference. The amplifier used was a Honeywell Biomedical Amplifier (Appendix C), with the low-frequency cutoff raised to 1 Hz to reduce baseline wandering. Gain settings varied from 5000 to 20 000 depending on the subject, and the amplifier was tied directly to the recorder.

This configuration was quite sensitive to even single spikes or minimal finger motion (Figure 12).

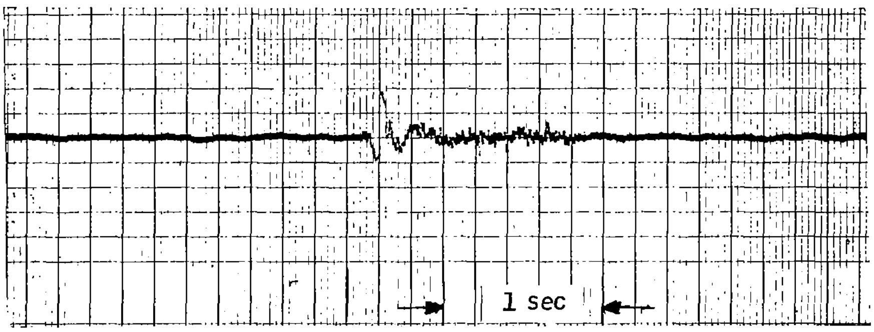


Figure 12. Sample Electromyogram Showing Minimal Finger Motion

The electrodes used for EMG and all monitoring except the ground electrode (right leg) were standard 1-cm silver E&M ECG electrodes affixed with adhesive washers and filled with Sanborn Redux electrode paste.

Respiration

Human respiration (more properly ventilation) is regulated in response to temperature, pneumotaxic, blood pO_2, pCO_2 and pH, and muscle stretch, as well as overriding voluntary control (Lim, 1966). Although the neuroanatomy of the system is well understood, there are many unanswered physiological questions, such as how the control parameters interrelate to determine rate and depth of ventilation.

There is a particular paucity of literature treating changes in the respiratory patterns with information workload.

From the pilot study, however, respiratory amplitude seemed very promising.

The subject's respiration was monitored by measuring the self-impedance change between two electrodes placed along the midaxillary line at the sixth intercostal space. The impedance pneumograph uses a 1- to 2-mA, 25-kHz exciting current, and the E&M physiograph was ac-coupled to the recorder. The frequency response of this system was 0. 1 to 100 Hz.

The respiration signal was particularly encouraging, since it exhibited visually detectable changes between tasks on some subjects. Figure 13 is the respiration signal taken before and after a step change in dynamics from K/S to K/S^2.

Vectorcardiogram

The electrocardiogram (ECG) is the surface manifestation of the superposition of currents during activation of individual heart muscle fibers. Thus the ECG gives information about the direction and velocity of the cardiac excitation wave (electrical versus mechanical activity).

Extrinsic factors regulating cardiac activity include neurological, hormonal, and fluid mechanical. Their effect is outlined in Appendix B (Spyker, 1970), but the important point is that gross changes (e. g., heart rate) and subtle changes (e. g., T wave depression) do occur in the ECG as a result of autonomic cardiac regulation.

The scaler ECG is felt to contain adequate information for many applications, but it cannot represent the three-dimensional excitation wave. The SVEC III system is usually considered to be more accurate projection of the X, Y, and Z potentials (Schaeffer, 1965), but the Frank system requires fewer electrodes.

To record the VCG, a slightly modified Frank lead system (Figure 14) (Frank, 1956) was used. The modification was the use of a unity-gain buffer amplifier to provide impedance matching. This is important because of the skin/paste and paste/electrode polarization artifacts which would otherwise cause unrealistically high resistances in the Frank method.

The orthogonal X, Y, and Z outputs were fed to differential amplifiers with gains of 100 and a measured common mode rejection of at least 160 dB. The amplified signals were recorded on a multichannel strip chart recorder and on an FM magnetic tape recorder. Figure 15 is a sample VCG record.

Skin Impedance

Benson, et al (1965), indicates that galvanic skin resistance represents the largest emotional response under workload conditions. However, considerable confusion can arise in measuring and reporting galvanic skin

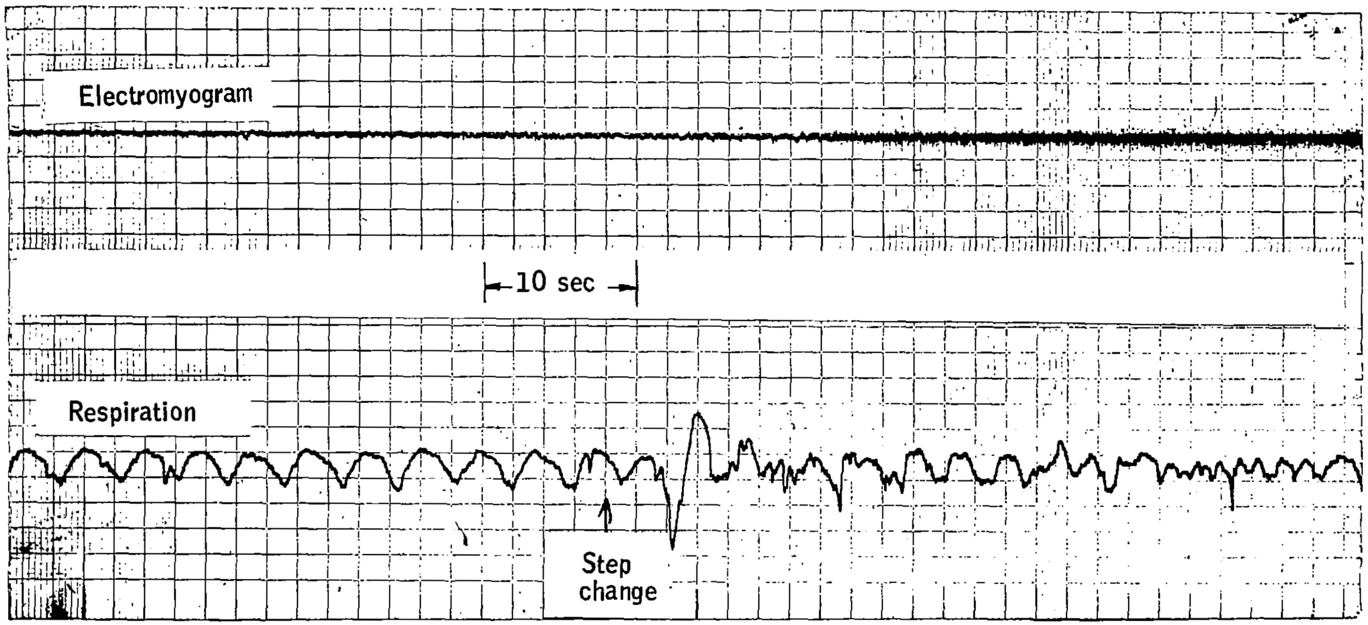


Figure 13. Respiration and Electromyogram During Step Change in Dynamics

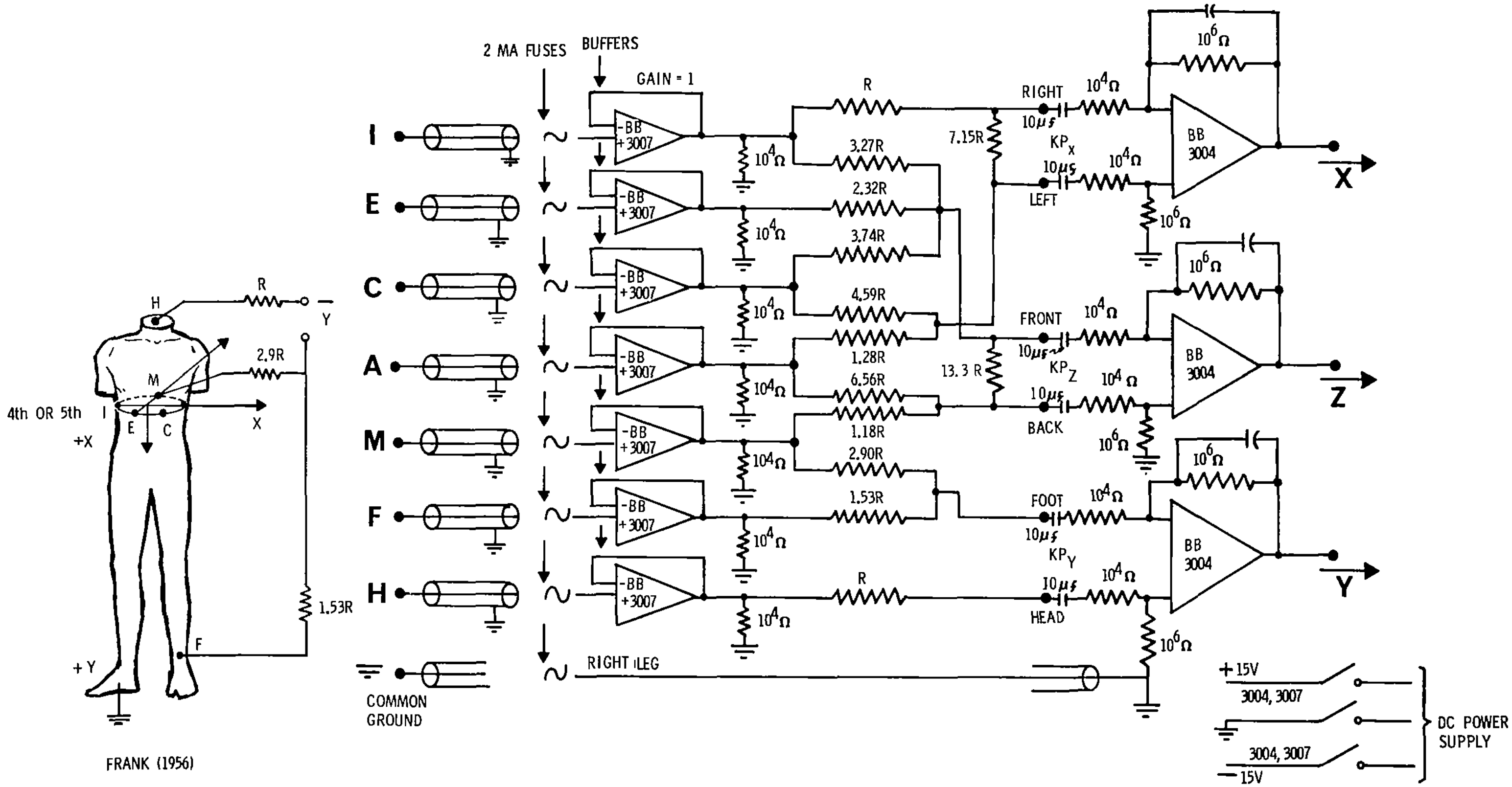


Figure 14. Space Vectorcardiographic Frank Lead System with Buffer Amplifiers in Pickup Electrodes; Gain from Weighting Network is 100 through X, Y, and Z Components (R = 10^4 ohms)

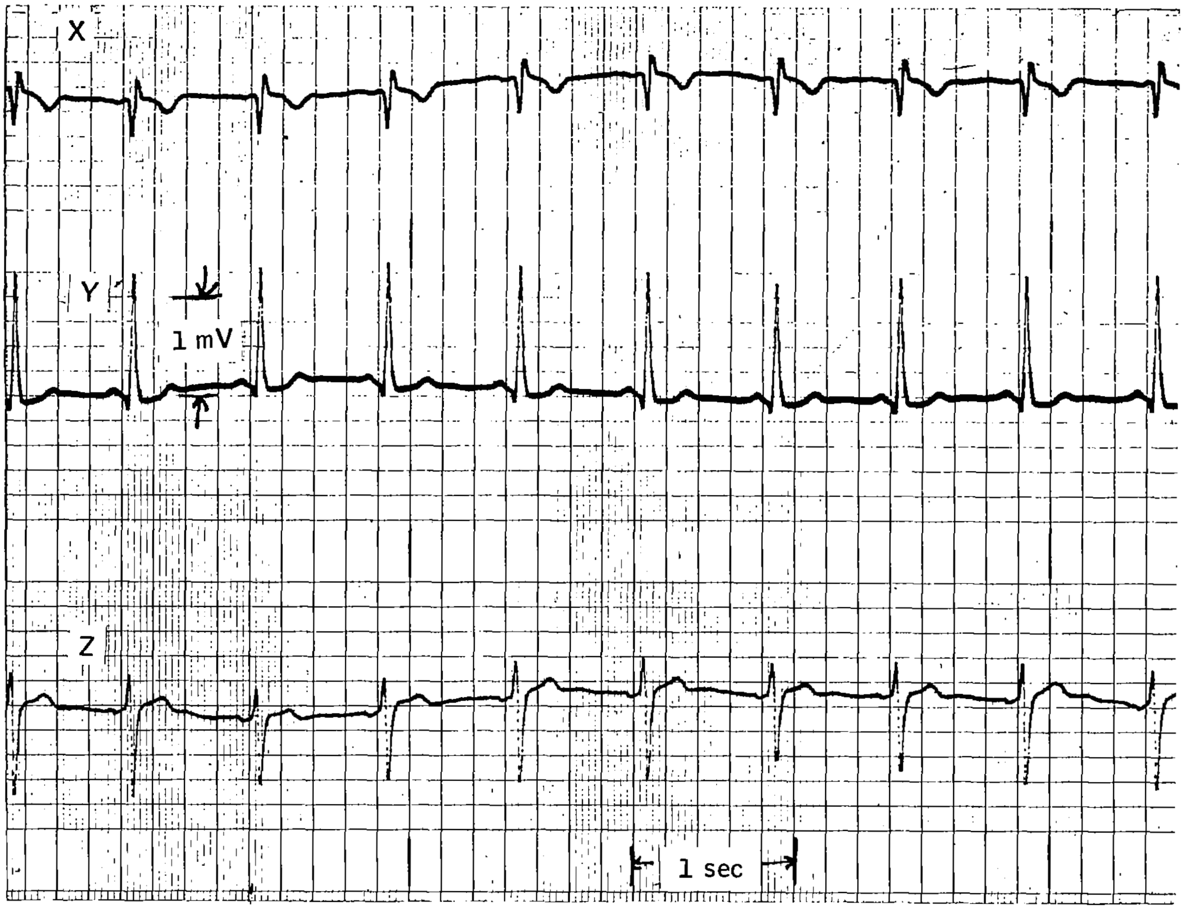


Figure 15. Sample of VCG Data from Frank Lead System

resistance. Reported values usually refer to the system skin-paste-electrode as a whole, without correcting for the skin-paste and electrode-paste polarization impedances. Previous work (Sinbel, 1966) indicates that the electrical characteristics of paste materials can profoundly affect the measured impedance value.

For these reasons we developed a hybrid system which permitted simultaneous measurement of the complex impedance and polarization for a subject + paste + electrode and for the electrode + paste (Figure 16). Preliminary investigations were carried out on-line, but during the main experiment the excitation current and voltage were recorded for off-line processing.

The volar self-impedance was measured between a standard 1-cm silver electrode on the sole of the right foot and the 1- x 1.2-inch German silver ground electrode (right ankle).

The excitation current (20 μA) and resulting voltages were recorded on two FM magnetic tape channels for off-line digital processing. The excitation was turned off until 3 minutes and 10 seconds into the 4-minute run when approximately 5 seconds of each of the following frequencies were recorded: 10, 20, 40, 80, 120, 170, 200, 400 and 800 Hz. The frequency, resistance, and reactance were then determined. A calibration signal with a 16-kΩ resistor was recorded at 120 Hz before each session.

Electroencephalogram

The electroencephalogram (EEG) represents a spatial average of neuroelectric activity which is remarkably similar in appearance to narrow-band gaussian noise. Our interest in the EEG for this study was limited to the visually-evoked brain response (VEBR) which is obtained by further time averaging of the EEG for 500 ms following a visual stimulus.

The stimulus consisted of a strobe lamp (General Radio Type 1531) (0.014 ft-c in the plane of the subject) placed 12 inches above the secondary task and was viewed binocularly at a distance of 3 feet. The stimulus was 10 μs in duration and was presented randomly with a mean intersignal interval of 3 seconds with a standard deviation of 0.5 second. The strobe lamp was synchronized with the onset of the secondary task lights (Figure 2). The EEG was amplified through a Honeywell Biomedical Amplifier at a gain of from 3 to 20 x 10^5. The output of the amplifier was fed into a filter (Krohnhite) which had a bandpass of 0.2 Hz to 40 Hz and then to the FM tape recorder.

The EEG was differentially taken from two needle-type electrodes inserted under the scalp, both along the midline, one at and the other 2 cm above the inion.

Figure 17 shows the EEG with characteristic alpha rhythm when the subject's eyes are closed.

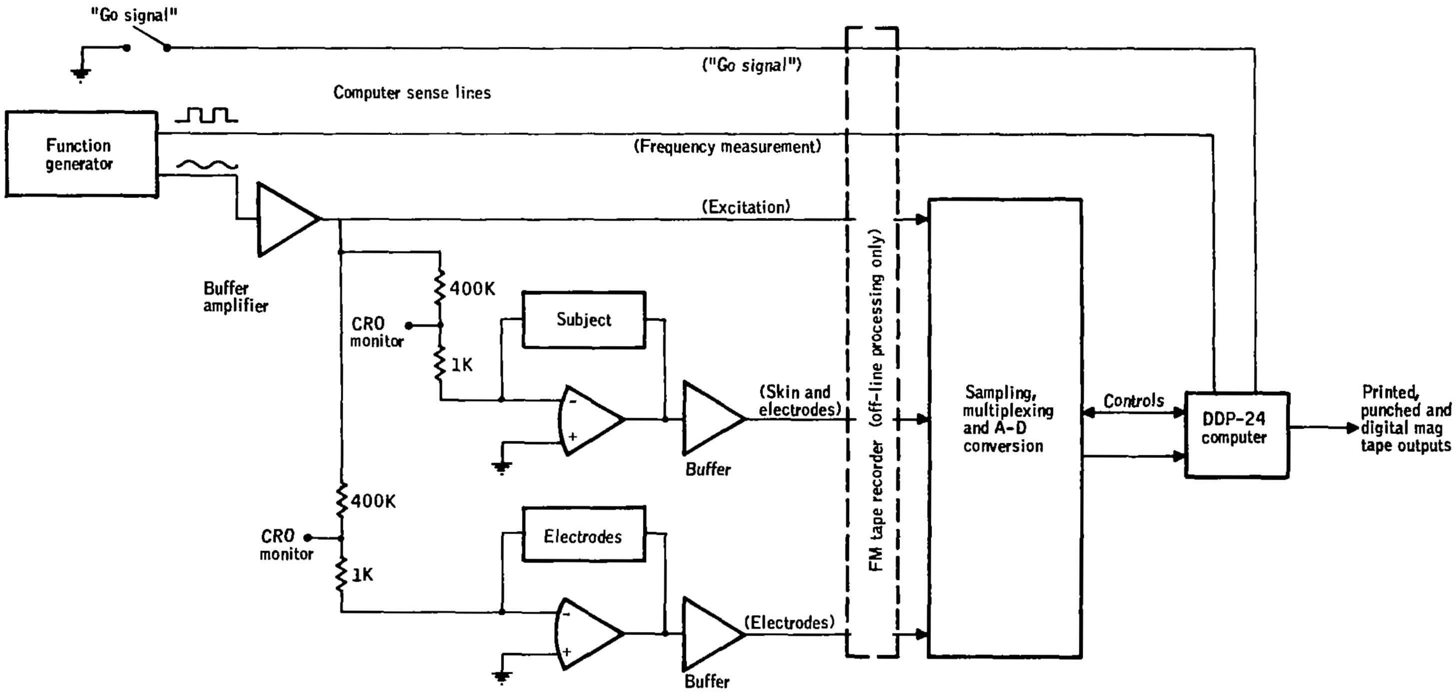


Figure 16. Schematic of Skin Impedance Measurement System

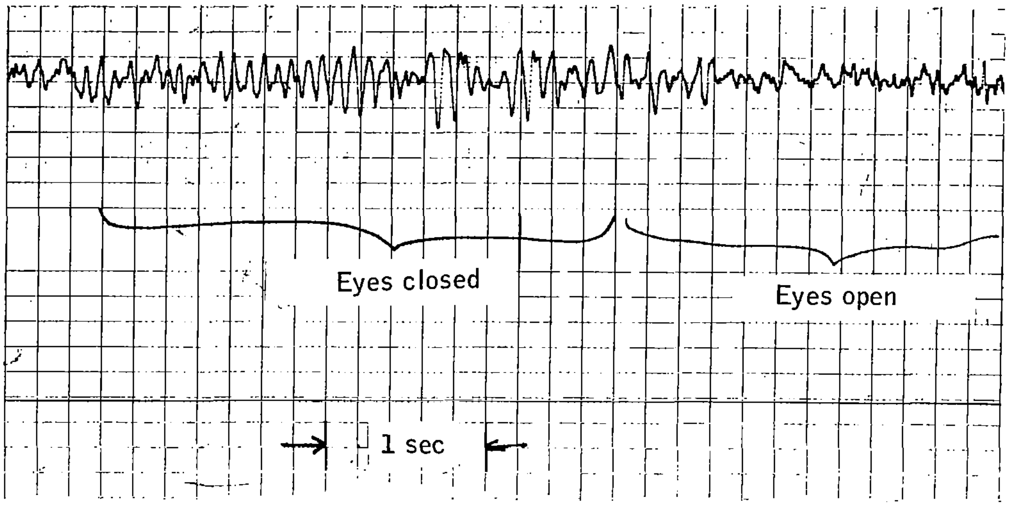


Figure 17. Sample EEG Showing Good Alpha Activity

Figure 18 summarizes the physiological monitoring system and illustrates the grounding system which permitted good quality miltichannel data collection. Unshielded electrode wires were kept as short as possible and went only to the chair-mounted electrode panel. All other leads were individually shielded. The only subject ground was the right ankle electrode.

FEATURE EXTRACTION

In the preceding section, the creation of the analog data base for this study was outlined. The remainder of the report will be concerned with the steps in the transformation of this data base to a meaningful workload index predictor. This section will detail the first of these steps which include:

1) Selection and extraction of features

2) Normalization of data

3) Selection of a subset of "best" features

4) Simultaneous solution for the best (least-square) workload index and its linear predictors.

A basic premise in this study is that each 4-minute simulation with its particular noise cutoff and dynamics represents a discrete workload. Thus, if we

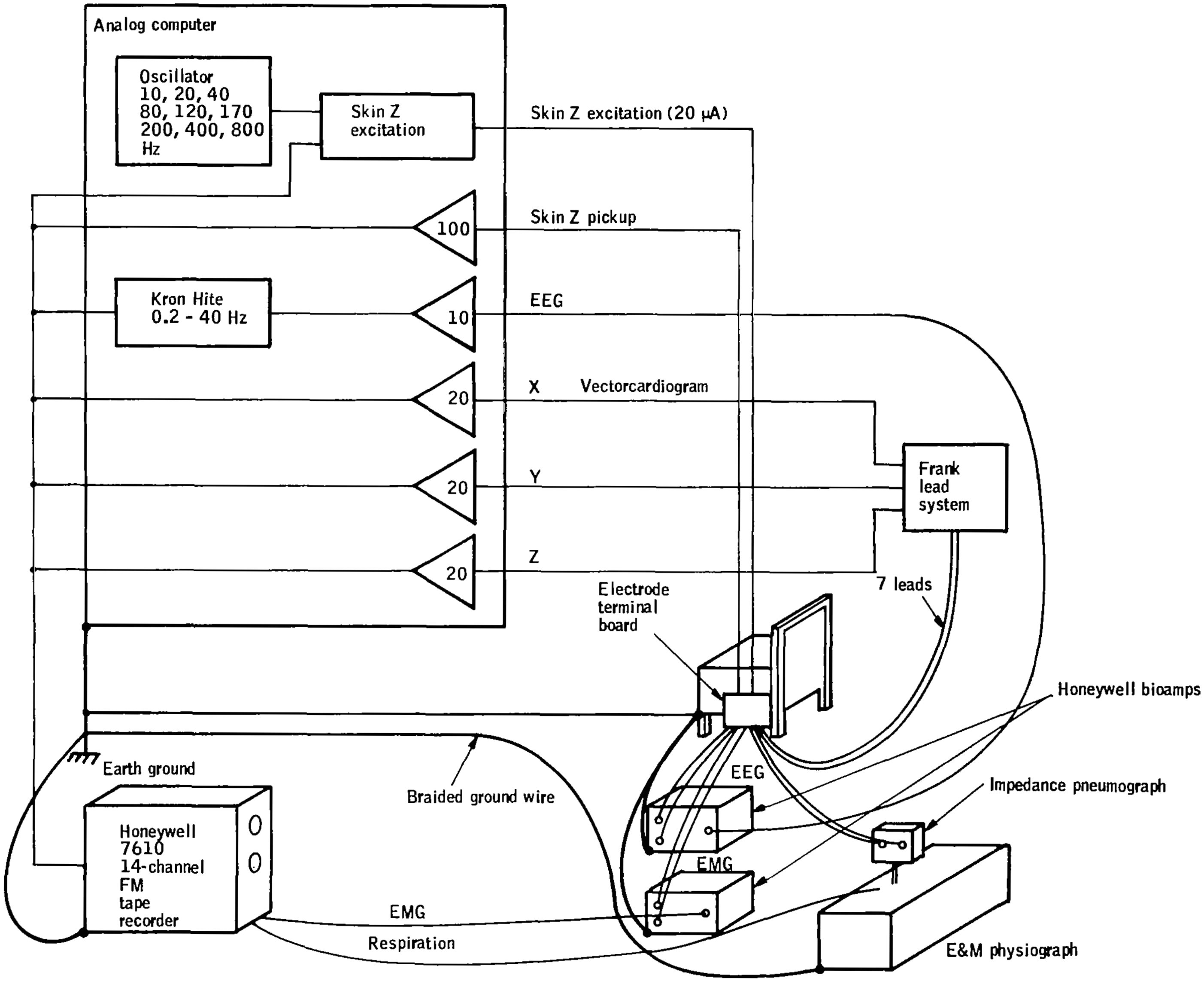


Figure 18. Physiological Monitoring System Summary

seek to predict this workload based on the physiological data, then the latter must be also characterized in terms of discrete variables (features).

Features may be chosen in a particular situation: a) as a result of insight into the underlying mechanisms, or b) by the try-everything (brute force) technique. In most real-life situations, the choices lie somewhere between these extremes and are often influenced by direct observation. For example, it was observed that the respiration seemed to increase in amplitude and decrease in regularity as a result of easy to hard changes in task. Thus, mean amplitude and variance in peak-to-peak interval were considered as good candidates for respiration features.

In our approach to workload prediction, features were selected to characterize the physiological steady state. We believe that a predictor's utility would be seriously limited if it were based on the early transient changes in physiological variables.

Electromyogram, respiration, and VCG features are all extracted from minutes 2 and 3 of the 4-minute run. Visually-evoked response is averaged from the 10^{th} through 60^{th} stimuli of the approximately 80 which occurred in each run and the skin impedance measurements (9 excitation frequencies) were taken serially at the beginning of minute 4.

Table III lists the 84 physiological features which were extracted for each 4-minute session, and the following discussion describes the procedure used. In all cases, the feature extraction was carried out automatically on either the analog or digital computer and in most cases at 8x real time. The data were visually monitored for gross artifact during sampling and in some cases on the computer-driven CRT during the extraction process (Figure 19).

Electromyogram

The off-side (left arm) electromyogram was bandpass filtered to reduce influences of baseline wandering, squared, and integrated on the analog computer (Figure 20). The square root of this value was used as the EMG feature:

$$\text{IEMG} = \frac{1}{2}\left[\int_1^3 s(t)^2 \, dt\right]^{1/2}$$

The mean values, by task, for the integrated EMG and all physiological features are included in Appendix D.

Respiration

Visual examination of the respiration data suggested that in some subjects amplitude and regularity were affected by workload. Thus, although some preliminary work was done with power spectral density (Fast Fourier

TABLE III. - PHYSIOLOGICAL FEATURES

Feature number	Feature description
5	Integrated electromyogram
	Respiration features
6	Mean amplitude, low
7	S.D. amplitude, low
8	Mean amplitude, high
9	S.D. amplitude, high
10	Mean interval, low
11	S. D. interval, low
12	Mean interval, high
13	S.D. interval, high
14	Signal average, low
15	Signal power, low
16	Signal average, high
17	Signal power, high
18	Rectification, low
19	S.D. rectification pieces, low
20	Rectification, high
21	S.D. rectification pieces, high
	Vectorcardiogram features
22	R-wave amplitude, mean (mV)
23	R-wave amplitude, S.D. (mV)
24	S-T amplitude, mean (mV)
25	S-T amplitude, S.D. (mV)
26	T-wave amplitude, mean (mV)
27	T-wave amplitude, S.D. (mV)
28	Baseline T-P, mean (mV)
29	Baseline T-P, S.D. (mV)
30	R-T interval, mean (seconds)
31	R-T interval, S.D. (seconds)
32	R-R interval, mean (seconds)
33	R-R interval, S.D. (seconds)
	Skin impedance features
34	Series resistance, R_s, (kΩ)
35	Parallel resistance, R_p (kΩ)
36	Leakage conductance, G (μmhos x 100)
37	Capacitance, C (μF x 100)
38	Cord angle, phi (deg)
39	Average radius (kΩ)
40	Standard deviation of error (kΩ)
41	Circle center, real (kΩ)
42	Circle center, imaginary (kΩ)
43	10 Hz skin Z, real
44	20 Hz skin Z, real

TABLE III. - PHYSIOLOGICAL FEATURES - Concluded

Feature number	Feature description
45	40 Hz skin Z, real
46	80 Hz skin Z, real
47	120 Hz skin Z, real
48	170 Hz skin Z, real
49	200 Hz skin Z, real
50	400 Hz skin Z, real
51	800 Hz skin Z, real
52	10 Hz skin Z, reactive
53	20 Hz skin Z, reactive
54	40 Hz skin Z, reactive
55	80 Hz skin Z, reactive
56	120 Hz skin Z, reactive
57	170 Hz skin Z, reactive
58	200 Hz skin Z, reactive
59	400 Hz skin Z, reactive
60	800 Hz skin Z, reactive
	Visually-evoked response features
61	Signal power (μV)
62	Overall maximum (μV)
63	Latency overall maximum (ms)
64	Overall minimum (μV)
65	Latency overall minimum (ms)
66	Minimum 100 to 160 (μV)
67	Latency min. 100 to 160 (ms)
68	Maximum 150 to 220 (μV)
69	Latency Max. 150 to 220 (ms)
70	Minimum 180 to 290 (μV)
71	Latency min. 180 to 290 (ms)
72	Maximum 215 to 270 (μV)
73	Latency max. 215 to 270 (ms)
74	Sequential min. 1 (μV)
75	Latency min. 1 (ms)
76	Sequential max. 1 (μV)
77	Latency max. 1 (ms)
78	Sequential min. 2 (μV)
79	Latency min. 2 (ms)
80	Sequential max. 3 (μV)
81	Latency max. 2 (ms)
82	Sequential min. 3 (μV)
83	Latency min. 3(ms)
84	Sequential max. 3 (μV)
85	Latency max. 3 (ms)
86	Sequential min. 4 (μV)
87	Latency min. 4 (ms)
88	Number of maximums

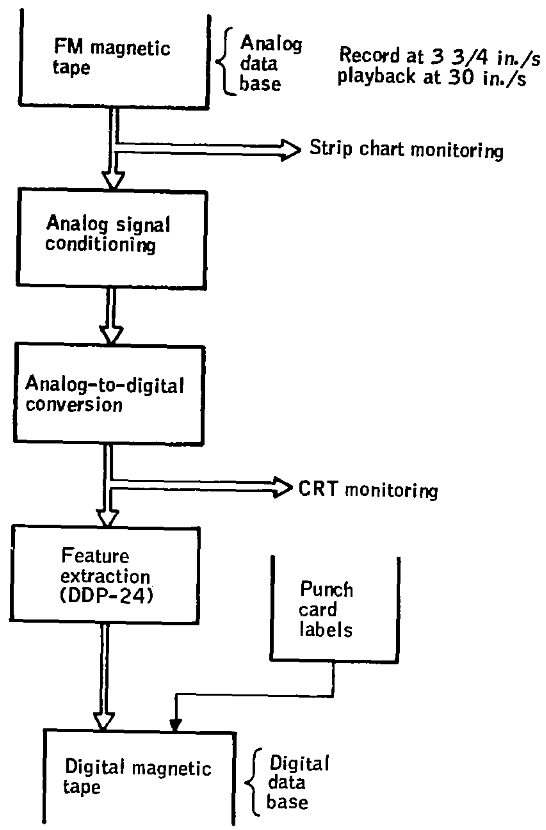


Figure 19. Automatic Feature Selection

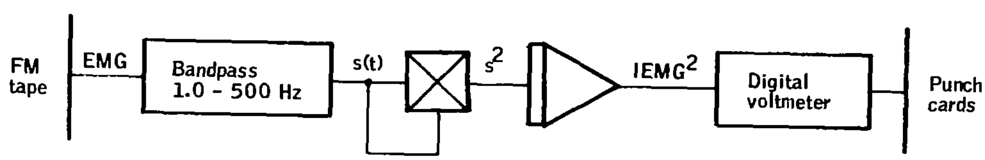


Figure 20. Processing for Integrated EMG

Transform), the features which were finally chosen were all basically means and variances of the amplitudes and intervals.

Figure 21 shows a relatively poor respiration signal and presents the motivation for the use of the two filters. In addition to the necessary smoothing, they may be said to provide two distinct definitions of "breath". The normal respiratory rate is 20 breaths/minute (0.3 Hz). So the second-order low-pass filter (0 to 0.14 Hz) is only responsive to large amplitude, slow respiration. The bandpass filter (0.125 to 1.3 Hz) preserves most of the waveform detail.

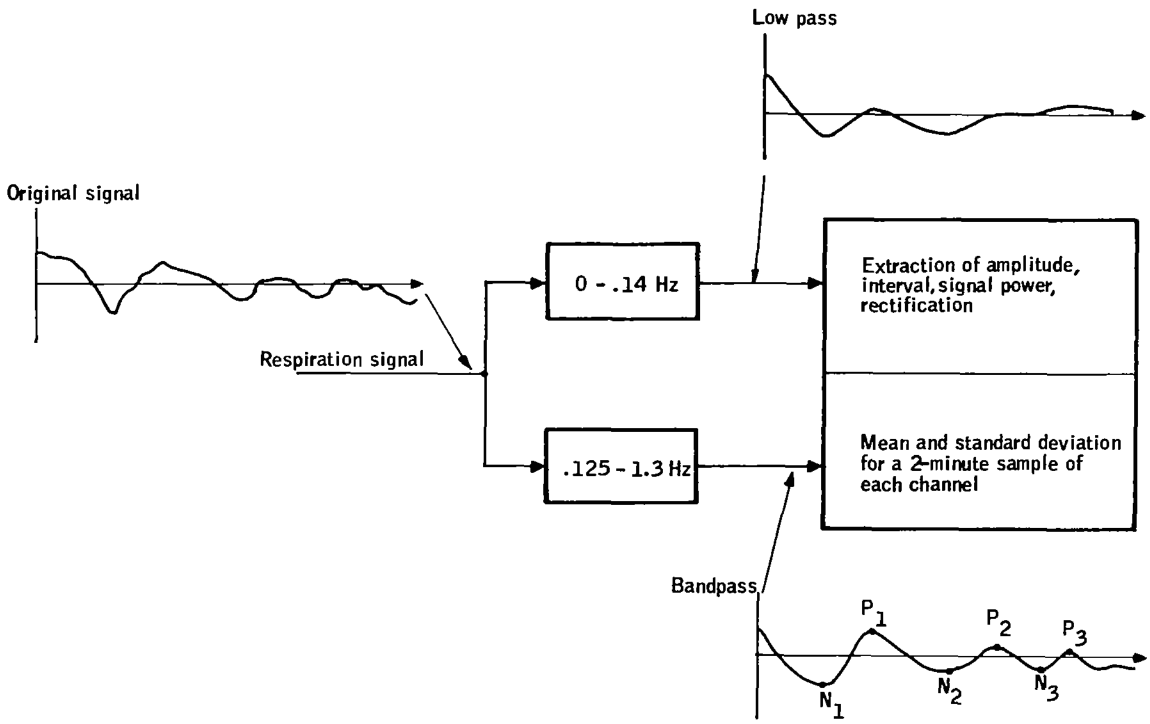


Figure 21. Respiration Processing

These two channels were sampled simultaneously at 10 samples per second and then smoothed with a moving average vector which was 5 samples wide (zero-order curve fit). This was necessary, since a zero derivative was used to define maximum and minimum points and sampling easily introduces multiple peak errors. The changes introduced by the smoothing were visually undetectable.

The maxima (P_i) and minima (N_i) were then found, and the following eight features computed for each channel:

	Feature Number Low pass	Feature Number Bandpass
1) Average amplitude: $\bar{A} = \frac{1}{N} \sum_{i=1}^{N} (P_i - N_i)$	(6)	(8)
2) Standard deviation of amplitude: σ_A	(7)	(9)
3) Average interval: $\bar{I} = \frac{1}{N} \sum_{i=1}^{N} (P_{i+1} - P_i)$	(10)	(12)
4) Standard deviation of interval: σ_I	(11)	(13)
5) Signal average: $P_o = 1/2 \int_1^3 s(t)\, dt$	(13)	(16)
6) Signal power: $P = 1/2 \left[\int_1^3 s(t)^2\, dt \right]^{1/2}$	(15)	(17)
7) Signal rectification: $R = \sum_{i=1}^{N} \Big(\lvert P_i - N_i \rvert + \lvert N_i - P_{i+1} \rvert \Big)$	(18)	(20)
8) Standard deviation: σ_R, where $R_i \triangleq \Big(\lvert P_i - N_i \rvert + \lvert N_i - P_{i+1} \rvert \Big)$	(19)	(21)

Amplitude and interval are self-explanatory. Signal average was only a check. Since the waveform was ac-coupled to the recorder, it was always near zero. Rectification was used to provide a measure of total ventilation. This could be approximated by amplitude/interval, but it was decided not to use any nonlinear feature processing. The standard deviation of these pieces (8) is essentially the same as σ_A and was included mostly for programming convenience.

The proper labels (subject and run number) for the 16 features were written onto digital magnetic tape by simply reading them off a punch card. Once these labels were punched for each FM tape, the process was virtually automatic.

Electrocardiogram

The VCG was handled in the same manner as the respiration signal. The magnitude of the vectorcardiogram $[X^2 + Y^2 + Z^2]^{1/2}$ was found to supress the S- and T-waves and was thus abondoned in favor of the X-channel only. This signal was filtered (0.1 to 40 Hz), amplified to ensure an R-wave trigger, and sampled for 2 minutes at the rate of 100 samples per second.

The peak of each R-wave was located, and the following determinations were made (Figure 22):

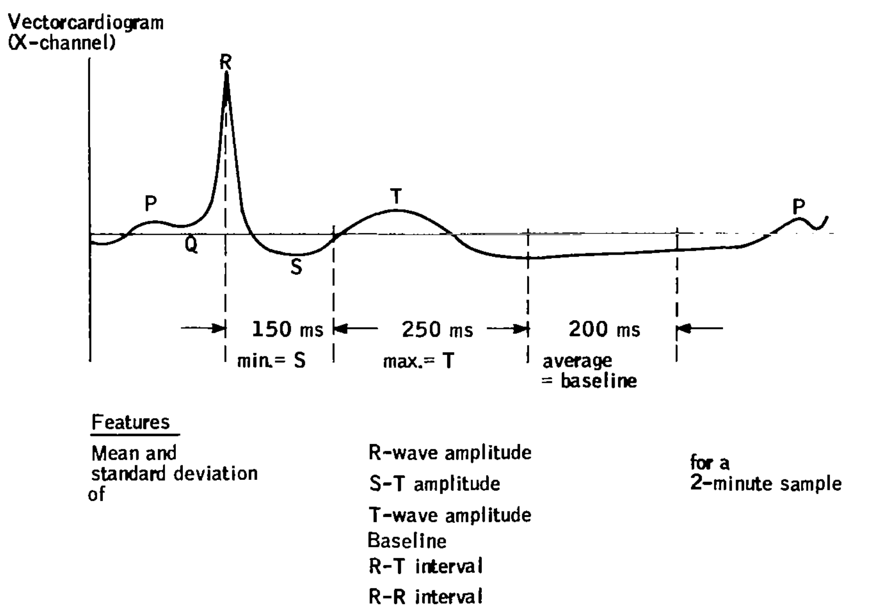


Figure 22. Electrocardiogram Features

- S: Minimum over next 150 ms post-R peak
- T: Maximum over next 400 ms post-R peak
- Baseline: Average of 400 to 600 ms post-R peak

The features which were then recorded on digital magnetic tape are the mean and standard deviation over the 2-minute sample:

		Feature No.
1)	Mean R-wave amplitude (above baseline	(22)
2)	σ_R amplitude	(23)
3)	Mean S-T amplitude	(24)
4)	σ_{S-T} amplitude	(25)
5)	Mean T-wave amplitude (above baseline)	(26)
6)	σ_T amplitude	(27)
7)	Mean of baseline (from zero) (represents area of R and T waves)	(28)
8)	σ baseline	(29)
9)	Mean R-T interval	(30)
10)	σ_{R-T} interval	(31)
11)	Mean R-R interval	(32)
12)	σ_{R-R} interval	(33)

The feature descriptions are largely self-explanatory, but it should be pointed out that since the signal is zero average (ac-coupled), the baseline voltage (5) represents the area of the R and T waves.

Skin Impedance

Galvanic skin response (GSR) is the term used to describe the small, rapid (seconds) change in skin resistance following stimulus. It is reportedly the most sensitive physiological indicator of psychological events available. The slower (minutes) but significant changes in skin resistance is called the basel resistance level (BRL) and has been shown to provide a meaningful indicator of the subject's alertness (Levy, 1958). It is this basal skin impedance which was examined as a correlate of workload.

A plot of the resistive and reactive components of the skin as a function of frequency (Argand Plot) typically describes a semicircular plot with depressed center (Figure 23). A four-element model was recently proposed which neatly describes this characteristic (Khalafalla, 1970). The model consists of a series resistor, R_s, and a second fixed resistor, R_p, which is paralleled with an RC combination (Figure 24). It is assumed that the conductance G varies directly with frequency as does the capacitor, so that the impedance of the parallel branch is

$$Z_p = \frac{1}{G\omega} \parallel \frac{1}{jC\omega} = \frac{1}{\omega(G+jC)}$$

and of the whole model is

$$Z(\omega) = R_s + \frac{R_p}{1 + \omega R_p G + j\omega R_p C}$$

which describes a circular arc as ω goes from zero to infinity.

Thus, in addition to using the impedance measurements themselves as features, the model parameters were also derived and used. Measurements made at three frequencies are sufficient to characterize the semicircle and hence the model, but serial measurements were made at nine frequencies from 10 to 800 Hz to provide redundancies and assess goodness of fit of the model. This procedure is summarized in Figure 25.

The features which were recorded on digital magnetic tape were:

		Feature No.	
R_{10}	X_{10}	(43)	(52)
R_{20}	X_{20}	(44)	(52)
R_{40}	X_{40}	(45)	(54)
R_{80}	X_{80}	(46)	(55)
R_{170}	X_{170}	(47)	(56)
R_{200}	X_{200}	(48)	(57)
R_{400}	X_{400}	(49)	(59)
R_{800}	X_{800}	(51)	(60)

For the least-squares fit circle:

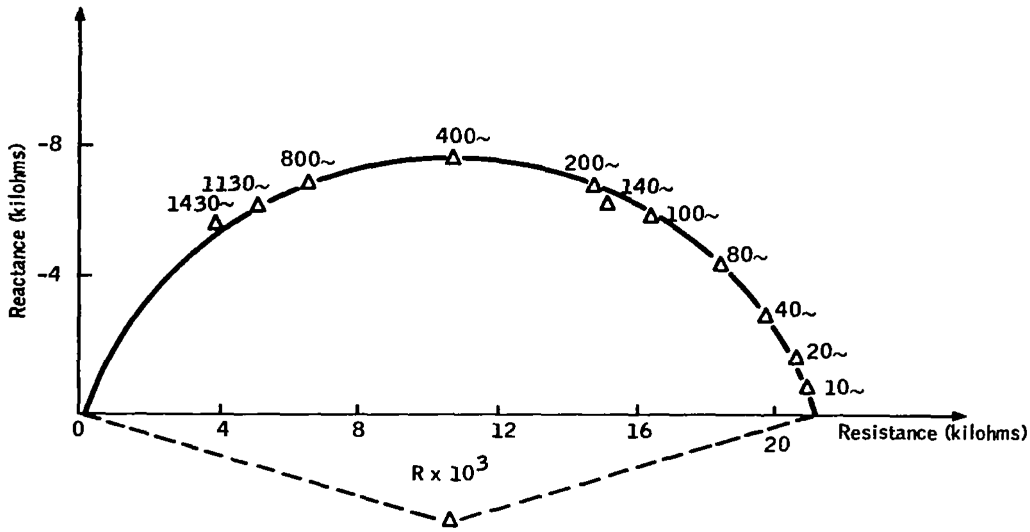


Figure 23. Representative Skin Impedance Arc Plot

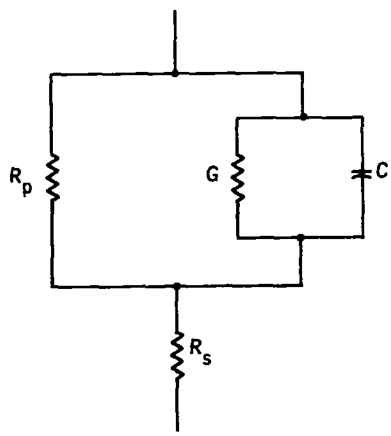


Figure 24. Electrical Model of Electrode Skin Impedance

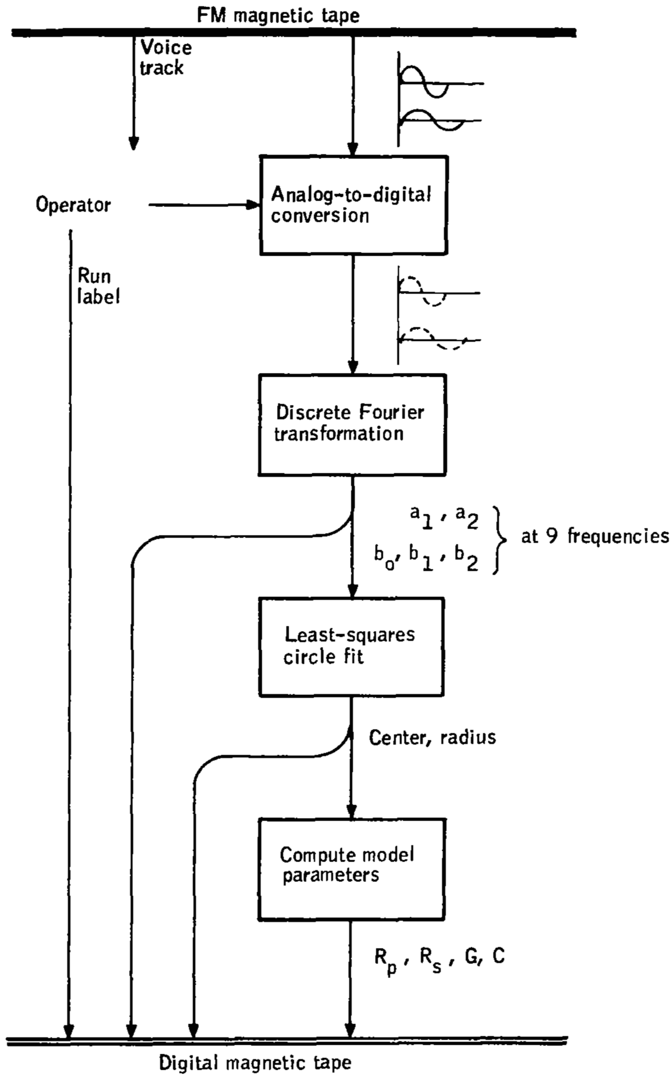


Figure 25. Summary of Skin Impedance Feature Extraction

Feature No.

- Average radius, $\bar{r}$ (39)
- Standard deviation of error,

$$\sigma_r = \sqrt{1/9 \sum_{i=1}^{9} (r_i - \bar{r})^2} \qquad (40)$$

- Circle center - real, R_{center} (41)
- Circle center - imaginary, X_{center} (42)

and the model parameters

- Series resistance,

$$R_s = X_{center} - \sqrt{\bar{r}^2 - X_{center}^2} \qquad (34)$$

- Parallel resistance,

$$R_p = -R_s + R_{center} + \sqrt{\bar{r}^2 - X_{center}^2} \qquad (35)$$

- Leakage conductance,

$$G = \frac{R_p (R_{120} - R_s) - (R_{120} - R_s)^2 - X_{120}^2}{[(R_{120} - R_s)^2 + X_{120}^2] R_p X_{120}}$$

- Membrane capacitance,

$$C = \frac{-X_{120}}{2\pi\, 120 \,[(R_{120} - R_s)^2 + X_{120}]} \qquad (37)$$

The circular arc fits were remarkably good:

- Average radius = 16 000 Ω
- Average error = 250 Ω

Figure 26 shows three plots with measured and model data.

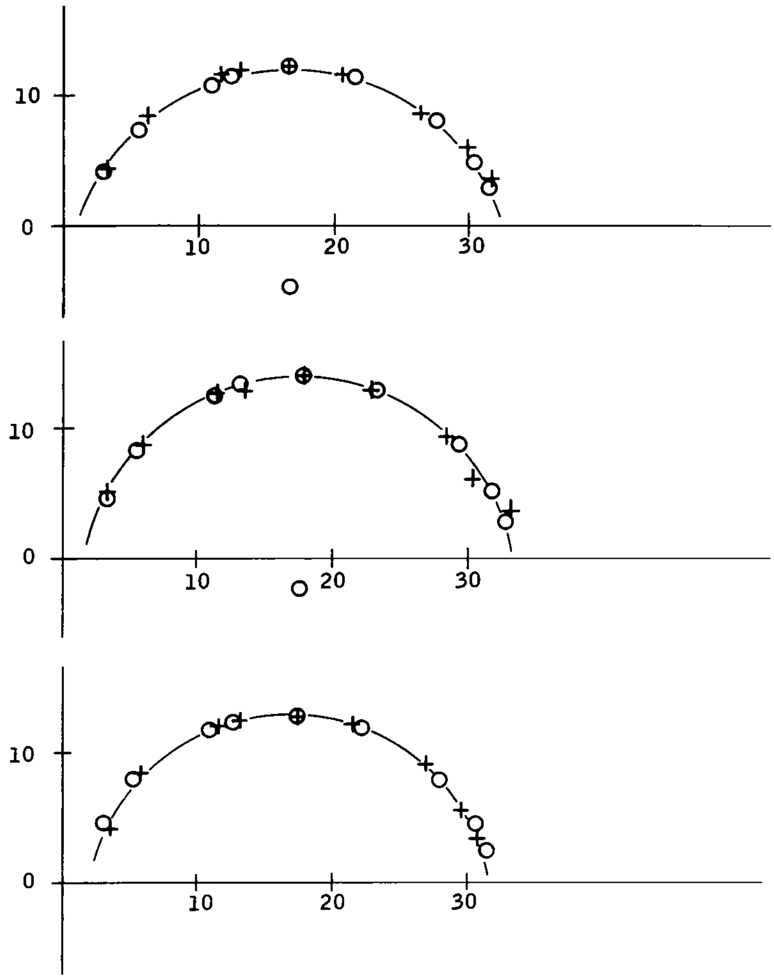


Figure 26. Measured (+) and Model Fit (0) Skin Impedance Data

Visually-Evoked Response

The time averaging of the EEG was done off-line on a special-purpose digital averager (Computer of Average Transients - CAT 400B). The pulse from the strobe light was recorded on the data tape channel adjacent to the EEG and used to trigger the CAT. The analysis period was 500 ms immediately following the stimulus, and 50 consecutive stimuli commencing 30 seconds into the run were averaged. The averaging could be performed at 30 in. / s (8x real-time) without detectable degradation of waveform. The CAT output (400 data points) was punched on paper tape for computer processing.

For the digital computer feature extraction the average of the first 20 samples (25 ms) was taken as zero, and all amplitudes were measured with respect to that reference. Since zero derivatives were used to define extrema, the signal was smoothed with a second-order, maximally-flat, low-pass digital filter with a cutoff at 30 Hz.

There are many ways to characterize a waveform of this type, but again it was decided to stick with simple time domain features. This was done in an attempt to automate the visual process of selecting and labeling peaks in the conventional manner (Figure 27). It would not, for example, be acceptable to merely consider successive maxima and minima, starting from either end, as features. The system arrived at was to partition the time axis and define the amplitudes and latencies based on the local extrema. The partitions were defined carefully, even though arbitrarily, after examining the X-Y plots of all EBRs.

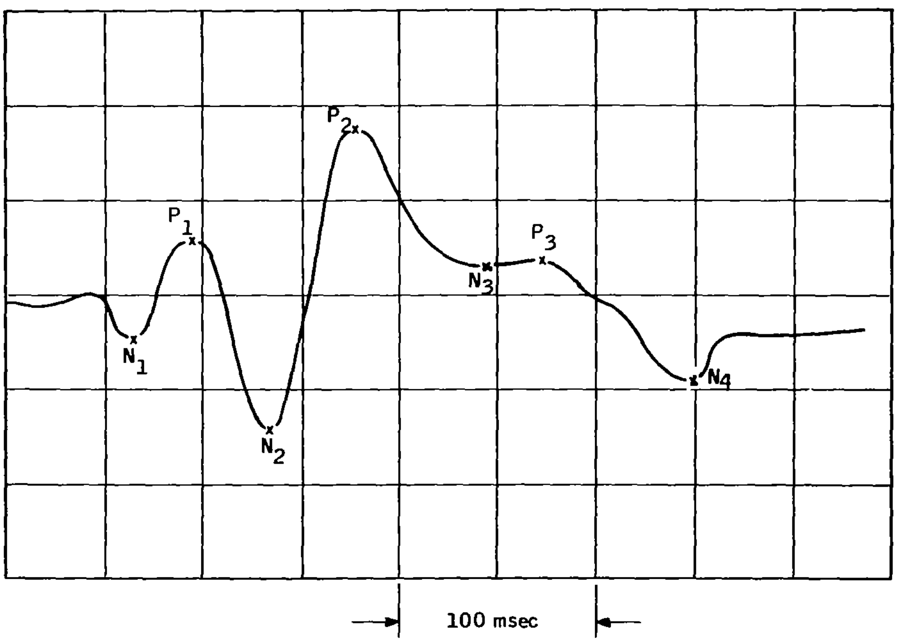


Figure 27. Sample of Visually-Evoked Waveform

The features extracted were:

• Signal power	(61)		
• Max. 0 to 400	(62)	Latency	(63)
• Min. 0 to 400	(64)	Latency	(65)
• Min. 100 to 160	(66)	Latency	(67)
• Max. 150 to 220	(68)	Latency	(69)
• Min. 180 to 290	(70)	Latency	(71)
• Max. 215 to 270	(72)	Latency	(73)

The max. 150 to 220 was chosen to label the P_2 wave which was a salient feature of most of the EBRs. An additional set of amplitudes and latencies (7 each) was then extracted which were sequential extrema in each direction from P_2 (feature number 80).

Figure 28 illustrates some representative VER data with the partitions.

FEATURE SELECTION

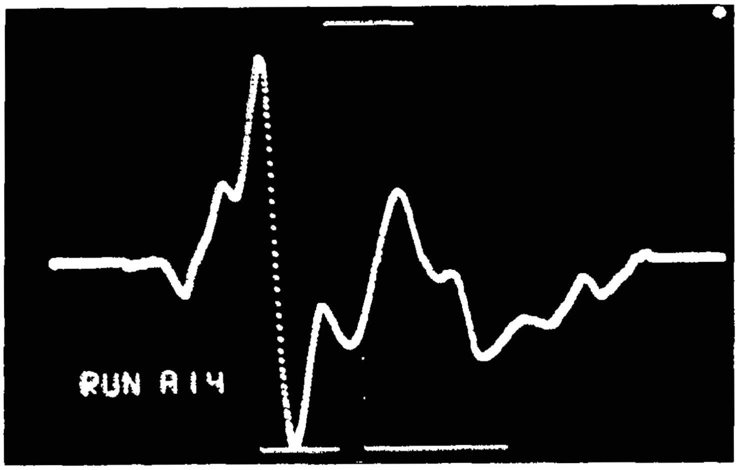
The previous sections have described the feature extraction for each physiological variable which resulted in the creation of digital data tape for respiration, vectorcardiogram, skin impedance, and evoked response. Tracking performance, secondary task performance, subjective rating and all physiological features were collected on a 90K magnetic drum. This included 243 runs from the main study, 60 runs from the validation study, and prebaseline physiological features.

The final objective is the prediction of the (as yet undefined) workload index based on the physiological features. Virtually any approach to this end will necessarily operate on a subset of these features. One of Honeywell's strengths in attacking this problem was an expertise in pattern recognition problems and the concomitant feature selection problems. Our experience during this study suggests that the discrete or classification approach is of limited value in selecting features for a linear predictor (which may be considered a continuous version of the classification problem).

In this section, the normalization procedure used for this study, results of feature selection using the classifier approach, and the multiple correlation approach which related more directly to the predictor development will be described.

Normalization

There are at least as many ways to approach the data analysis for this



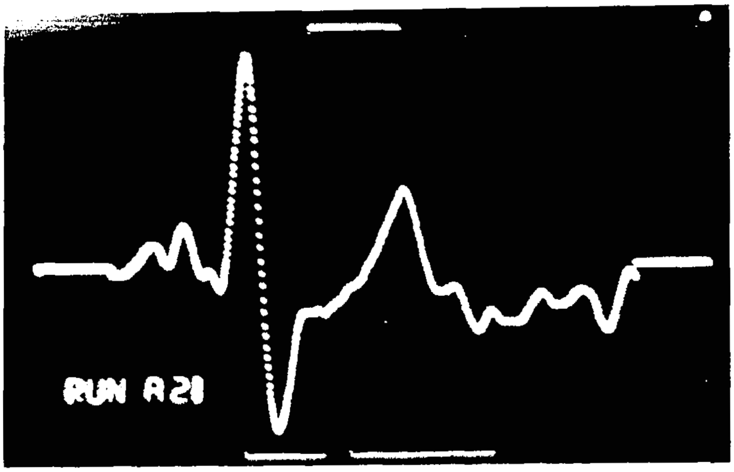


Figure 28. Representative VER Data with Partitions

problem as there are researchers. The one discussed here was chosen because it works and we feel it is defensible.

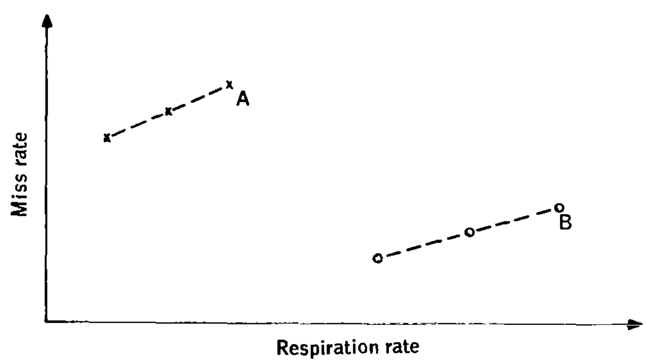
As mentioned in the preceding discussion, we are more concerned with measurement of relative workload. Thus a baseline value is needed from which to measure the change in the physiological features. The approach which proved to be the best for this study was to find the average of each feature for the nine consecutive runs in each session and use that as baseline. The normalized feature was then computed as percentage change from baseline. This procedure was also applied to the criteria variables (tracking error, miss rate, response time, and subjective rating). For example,

$$MR_{i(\text{normalized})} = \frac{MR_i - \overline{MR}}{\sigma_{MR}} \times 100, \; i = 1, 9$$

where $\overline{MR}$ is the average for the nine consecutive runs in that session.

Use of unnormalized data and normalization to presession baseline values was also investigated. The latter approach suffers from large differences between tracking and resting behavior. In particular, several subjects tended to take two or three unusually large breaths each minute during baseline.

A simple illustration can make the case for using the normalization procedure which we chose. Consider a hypothetical plot of miss rate versus respiration rate for two subjects:



Although each individual's data is clearly positively correlated, the pooled result would be negative correlation. This is precisely the problem that was avoided using session average baseline.

The next step which preceded the correlation studies and most of the classifier work was to take the across-replication average. This final data base represents each of the nine pilot's performance data and physiological state on each of the nine tasks by an average of his three replications on that task. This set of 81 observations was the basis for all tables and figures throughout this report unless otherwise specified.

Feature Selection by Task Classification

The pattern recognition system which we applied to this feature set is a distribution-free, adaptive system which automatically performs cluster seeking, feature selection, and discriminant design under a single performance criterion (Wee, 1968; Wee, 1970). It is possible to use this system without supplying a priori information as to class membership, but almost without exception the results are substantially better when a training set is used.

Three approaches to feature selection by classification were considered:

1) Consider the data resulting from three selected (e. g., tasks 1, 4, 9) tasks as representing three classes and select the features which best distinguished these. In this case, the number of samples was (3 replications) x (9 subjects) x (3 tasks) = 81.

2) Based on primary and secondary task performance it seemed realistic to consider a two-class (easy to hard) division with tasks 1 through 6 in class 1 and tasks 7 through 9 in class 2.

Number of samples = (3 replications) x (9 subjects) x (9 tasks) = 243.

The results of these two approaches are summarized in Figure 29. Although the best separation achieved was 86.4 percent with combined data, it should be pointed out that this is with averaged and for the most part unnormalized data.

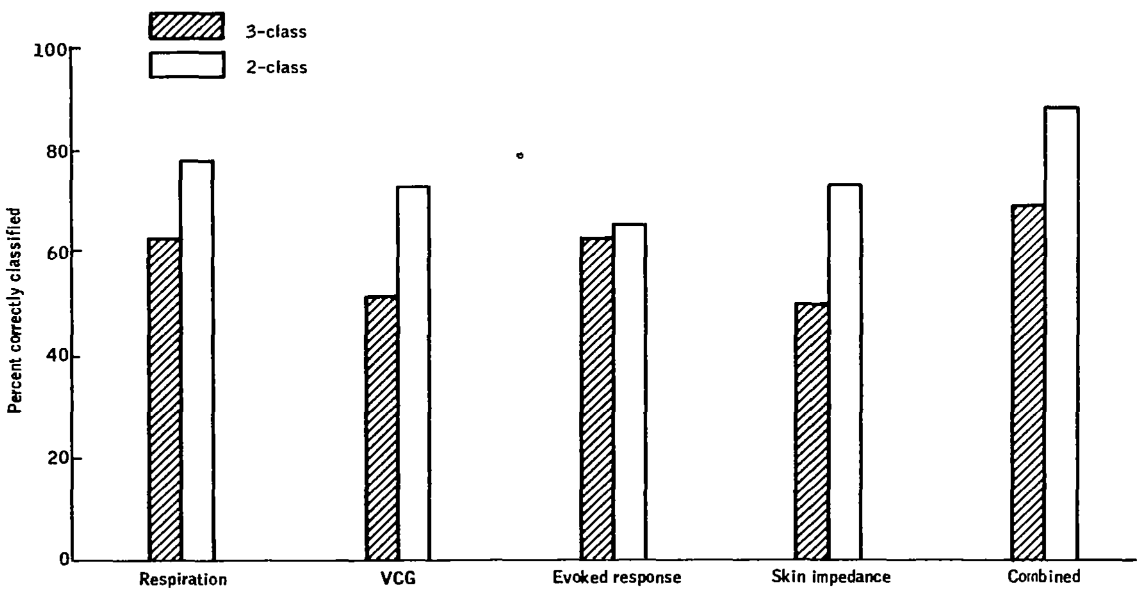


Figure 29. Summary of Preliminary Classification Results

Since it appeared that the data for the predictor would be best developed from the across-replication averages, the classifier was applied to the final data base assuming a two-class problem.

3) Class 1 = tasks 1 through 3

Class 2 = tasks 7 through 9

so the total number of samples was (9 subjects) x (6 tasks) = 54.

It may be well to point out at this juncture that there is a notable lack of significance tests for classifiers. The objective of such approaches is to achieve high-percentage separation on a large number of samples with a small number of features, but just what quantitative relation should exist between "high, large, and small" or how number of classes affects these has yet to be established. It is a "generally accepted" "rule of thumb" that for a two-class problem, the number of features (n) should not exceed the square root of the total number of samples (N). This has proved to be a very useful yardstick and we have observed:

1) The percentage separation achieved with any number of features is seldom substantially better than that achieved with $n = \sqrt{N}$.

2) It is not unusual to find percent recognition decreasing slightly when n gets greater than $\sqrt{N}$.

The classification results for combined features on the final data base are summarized in Figure 30. This figure illustrates the ordering of fea-

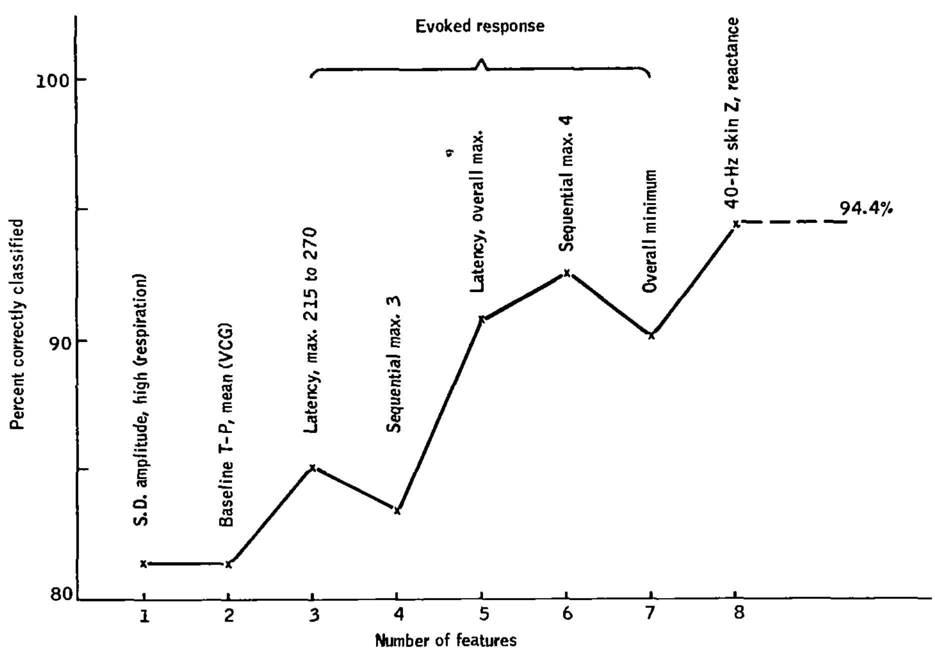


Figure 30. Classifier Performance - Combined Features

tures and percent separation for the main experiment (N = 54). The third through seventh features are from the evoked-response set. Since these results are not directly applicable to selection of features for the workload predictor, we will not discuss these results in any detail. It is worthy of mention, however, that equivalent results can be achieved using only respiration and VCG features (Figure 31).

Figure 32 presents the feature ordering for the validation study (N = 20) which achieved 100 percent separation with five features.

Multiple Correlation

The correlation coefficient

$$R = \frac{\overline{xy} - \bar{x}\bar{y}}{\sigma_x \sigma_y}$$

(where $\overline{}$ indicates mean or expected value)

is a measure of the degree of the linear relationship between two variables.

If x is a vector

$$x = \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix}$$

then the relation generalizes to

$$R_{ij} = \frac{\overline{X_i Y_i} - \bar{X}_i \bar{X}_j}{\sigma_{xi} \; \sigma_{xj}}$$

If we standardize the variables

$$X_i = \frac{X_i - \bar{X}_i}{\sigma_{xi}}$$

then the correlation matrix can be compactly written as

$$R_{xx} = \left[\overline{X X^T} \right] \qquad \text{(where } {}^T \text{ denotes transpose)}$$

which is symmetric with ones on the diagonal.

In the predictor development described in the next section the only information required will be the correlation matrix. The combined predictor will

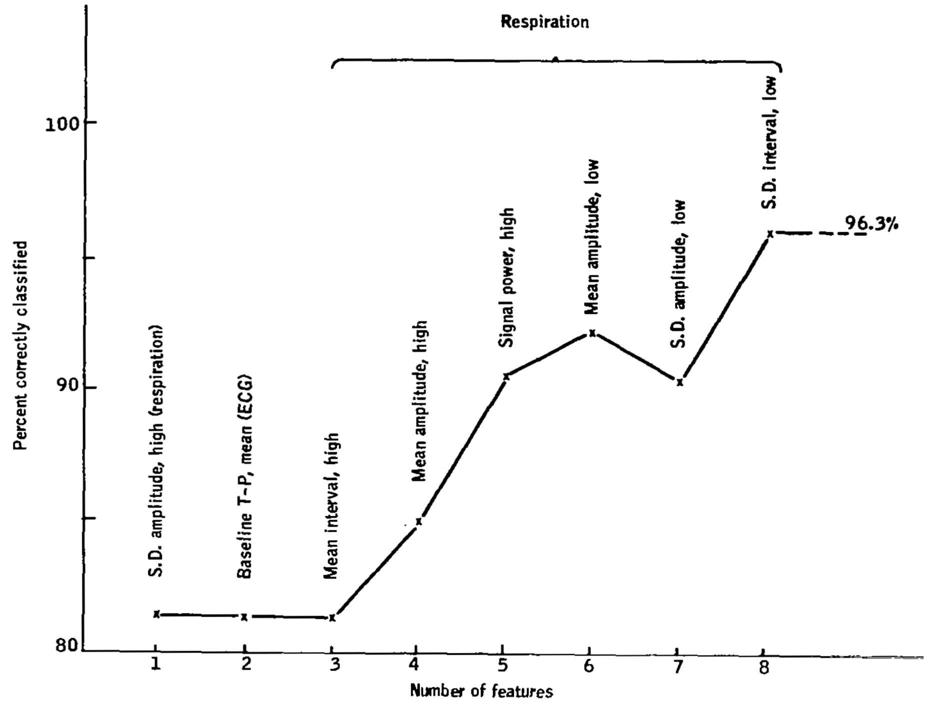


Figure 31. Classifier Performance - Respiration, VCG, and EMG Only

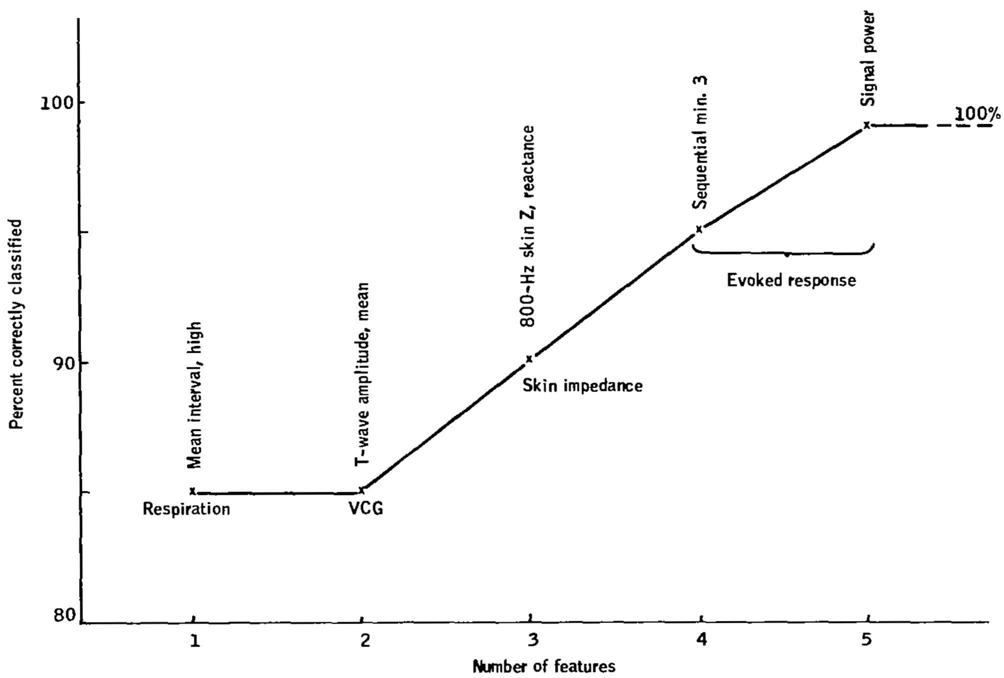


Figure 32. Classifier Performance - Validation Study Data

always be better than any correlation between an individual feature/criterion. For our purposes, criteria are defined as a measure of workload (or difficulty) as contrasted with features which are extracted from the physiological data and with which we will eventually predict the criteria.

The criteria of concern in this study are:

1) Tracking error (primary task performance)

2) Miss rate (error rate on secondary task)

3) Response time (average time to respond to lights)

4) Subjective rating (pilot's assessment of difficulty)

5) Task number(zero variance performance ranking)

6) Binary ranking (-1 for tasks 1 through 6, +1 for tasks 7 through 9)

These criteria were defined so as to be positively correlated (i.e., all generally increase with increasing workload) (Table IV).

TABLE IV. - CRITERIA CORRELATIONS

Feature	Tracking error	Miss rate	Response time	Subjective rating	Task number	Binary classification
Tracking error	1.000	.712	.686	.802	.848	.853
Miss rate		1.000	.760	.698	.637	.685
Response time			1.000	.717	.618	.648
Subjective rating				1.000	.893	.765
Task number					1.000	.897
Binary classification						1.000

Table V presents the correlation coefficients for all features with these six criteria.

TABLE V. - FEATURE/CRITERIA CORRELATION COEFFICIENTS

Feature number	Feature Description	Tracking error	Miss rate	Response time	Subjective rating	Task number	Binary ranking
5	Integrated electromyogram	.4946	.3719	.2861	.4353	.4990	.4838
	Respiration features						
6	Mean amplitude, low	.4537	.4528	.3122	.3794	.4530	.4638
7	S.D. amplitude, low	.5116	.4054	.2885	.4395	.5136	.5476
8	Mean amplitude, high	.3176	.3216	.1171	.2199	.3495	.3862
9	S.D. amplitude, high	.5867	.5002	.3056	.5298	.6036	.6785
10	Mean interval, low	.0704	.1045	.1632	.0938	.0523	.0000
11	S.D. interval, low	.4473	.3424	.2801	.4014	.4340	.4883
12	Mean interval, high	-.5767	-.4810	-.4417	-.6339	-.6256	-.5991
13	S.D. interval, high	.3592	.2382	.2035	.3431	.3759	.4320
15	Signal power, low	.4726	.4269	.3031	.3956	.4866	.5012
17	Signal power, high	.4788	.4639	.2698	.4221	.5236	.5527
18	Rectification, low	.4910	.4856	.3045	.4063	.5033	.5348
19	S.D. rectification pieces, low	.5118	.4055	.2886	.4391	.5133	.5475
20	Rectification, high	.5052	.4639	.2625	.4335	.5533	.5793
21	S.D. rectification pieces, high	.5866	.5000	.3055	.5293	.6031	.6770
	Vectorcardiogram features						
22	R-wave amplitude, mean (mV)	.1134	.0065	.0625	.0987	-.0011	.0579
23	R-wave amplitude, S.D. (mV)	.0533	-.0537	-.0128	-.0432	-.0618	.0123
24	S-T amplitude, mean (mV)	.3829	.2330	.2042	.3311	.2900	.3275
25	S-T amplitude, S.D. (mV)	.2961	.1948	.2123	.1824	.2642	.2634
26	T-wave amplitude, mean (mV)	.3460	.1891	.1693	.2820	.2425	.3234
27	T-wave amplitude, S.D. (mV)	.2622	.1998	.1983	.1296	.2163	.2745
28	Baseline T-P, mean (mV)	-.1562	-.0729	-.1040	-.3299	-.2930	-.1774
30	R-T interval, mean (seconds)	-.1141	-.0112	-.0126	-.0735	-.0564	-.1716
31	R-T interval, S.D. (seconds)	.2381	.1670	.2280	.2295	.2776	.1700
32	R-R interval, mean (seconds)	-.2433	-.2325	-.0881	-.2773	-.3039	-.2320
33	R-R interval, S.D. (seconds)	.2284	.1814	.1661	.0663	.1118	.1427
	Skin impedance features						
34	Series resistance, R_s (kΩ)	-.0462	-.0830	-.1134	-.0257	-.0167	-.1005
35	Parallel resistance, R_p (kΩ)	.0382	.0589	.0047	.0377	.0191	.0384
36	Leakage conductance, G (μmhos)	-.1875	.0098	-.0026	-.1168	-.1498	-.0947
37	Capacitance, C (μF x 100)	.0360	.0856	.0654	.0151	.0421	.0345
38	Cord angle, phi (deg)	.0749	.0621	-.0707	.0450	.1146	.0596
39	Average radius (kΩ)	.0520	.0576	.0270	.0340	.0148	.0397
40	Standard deviation of error (kΩ)	.1461	.0781	.1206	.0446	.0353	.0590
41	Circle center, real (kΩ)	.0261	.0556	-.0127	.0536	.0212	.0340
42	Circle center, imaginary (kΩ)	.0599	.0731	.1079	.0502	.0187	.0367
43	10 Hz skin Z, real	.0190	.0558	-.0201	.0660	.0207	.1436
44	20 Hz skin Z, real	-.0255	-.0163	-.0569	.0479	.0134	.0895
45	40 Hz skin Z, real	-.0168	-.0041	-.0613	.0401	-.0016	.0725

TABLE V. - FEATURE/CRITERIA CORRELATION COEFFICIENTS - Concluded

Feature number	Feature description	Tracking error	Miss rate	Response time	Subjective rating	Task number	Binary ranking
46	80 Hz skin Z, real	.0185	.0202	-.0186	.0656	.0208	.1053
47	120 Hz skin Z, real	.0042	.0003	-.0223	.0567	.0243	.0914
48	170 Hz skin Z, real	0.0190	.0116	-.0172	.0742	.0358	.1074
49	200 Hz skin Z, real	.0194	-.0040	-.0224	.0669	.0427	.1102
50	400 Hz skin Z, real	-.0195	-.0341	-.0381	.0239	-.0030	.0655
51	800 Hz skin Z, real	-.0175	.0042	-.0230	-.0148	-.0146	.0576
52	10 Hz skin Z, reactive	.1149	.0775	.0801	.1104	.0772	.1166
53	20 Hz skin Z, reactive	-.0697	.0107	-.0369	.0053	-.0206	-.0075
54	40 Hz skin Z, reactive	-.0671	-.0215	-.1033	-.0546	-.0530	-.0481
55	80 Hz skin Z, reactive	-.0051	.0072	-.0467	.0169	-.0233	.0340
56	120 Hz skin Z, reactive	.0334	.0377	-.0364	.0506	.0209	.0237
57	120 Hz skin Z, reactive	.0129	.0231	.0508	.0681	.0196	.0096
58	200 Hz skin Z, reactive	.0087	-.0216	-.0764	.0329	.0215	.0028
59	400 Hz skin Z, reactive	-.0696	-.1736	-.1565	-.0699	-.0707	-.1163
60	800 Hz skin Z, reactive	-.1036	-.0026	-.0346	-.0458	-.1363	-.1072
	Visually-evoked response features						
61	Signal power (µV)	-.0098	.0048	.0431	-.0419	-.0723	-.0032
62	Overall maximum (µV)	-.0558	-.0393	-.0093	-.0937	-.1079	-.0381
63	Latency overall max. (ms)	-.1955	-.1659	-.2129	-.2159	-.2151	-.1903
64	Overall minimum (µV)	.2197	.1487	.0733	.1442	.1924	.1819
65	Latency overall min. (ms)	.1416	.1594	.1705	.2149	.1786	.0870
66	Minimum 100 to 160 (µV)	-.0332	-.0549	-.0823	-.0018	-.0603	-.0376
67	Latency min. 100 to 160 (ms)	.0338	.0602	.0635	.0963	.0372	.0178
68	Maximum 150 to 220 (µV)	.1531	.0897	.1271	.0826	.0814	.1833
69	Latency max. 150 to 220 (ms)	-.1528	-.1156	-.1253	-.0941	-.1701	-.1740
70	Minimum 180 to 290 (µV)	.2434	.2623	.2249	.2809	.2424	.2888
71	Latency min. 180 to 290 (ms)	-.1419	-.2099	-.2947	-.2571	-.1758	-.0531
72	Maximum 215 to 270 (µV)	-.0399	-.0333	-.0555	-.0906	-.0139	.0609
73	Latency max. 215 to 270 (ms)	-.3191	-.2163	-.2531	-.3912	-.3008	-.3976
74	Sequential min. 1 (µV)	-.1378	-.0874	-.0762	-.0922	-.1024	-.2017
75	Latency min. 1 (ms)	-.1564	-.0149	-.0644	-.0635	-.0907	-.1025
76	Sequential max. 1 (µV)	-.1386	-.0817	-.0609	.1278	.1952	.2143
77	Latency max. 1 (ms)	-.1780	-.0436	-.1294	-.0823	-.1295	-.1458
78	Sequential min. 2 (µV)	-.0625	-.0172	-.0160	.0167	.0248	-.0604
79	Latency min. 2 (µV)	-.1882	-.0700	-.1428	-.1132	-.1911	-.1745
80	Sequential max. 3 (µV)	.1852	.1223	.1498	.1310	.1221	.2122
81	Latency max. 2 (ms)	-.1497	-.1203	-.1332	-.1023	-.1691	-.1674
82	Sequential min. 3 (µV)	.2624	.2624	.2554	.1557	.1132	.2464
84	Sequential max. 3 (µV)	.0178	-.0481	.0598	-.0062	.0460	.0453
85	Latency max. 3 (ms)	-.1033	-.0680	.0337	-.0297	-.1205	-.0879
86	Sequential min. 4 (µV)	.1846	.0579	-.0133	.1308	.1302	.1571
87	Latency min. 4 (ms)	-.0847	-.0040	.0830	-.0197	-.1154	-.0535
88	Number of maximums	.0956	.1896	.1674	.0729	.0983	.0546

Table V is more of a store of information than a display but several important aspects should be underscored. First, as to significance, Snedecor and Cochran (1967) give the test for null hypothesis ($R = 0$). For $N = 81$, it is rejected with $p > .95$ for $R > .217$ and with $p > .99$ for $R > .283$. Thus, table entries larger than .283 represent significant relations at the 1-percent level.

Respiration is clearly the strongest feature, with several coefficients greater than .5. Vectorcardiographic features exhibit some significant correlations, as does the EMG, but there is a notable lack of significant correlations among the skin impedance and evoked-response features.

It should also be pointed out that there is substantial redundance in some of the features, as verified by their nearly identical correlation coefficients, most notably:

- Features 6 and 18
- Features 7 and 19
- Features 9 and 21
- Features 15 and 17
- Features 25 and 27

Concluding comments on feature selection will be witheld until after the discussion of predictor development.

WORKLOAD INDEX

From the preceding discussion it is seen that there exist significant relationships between the physiological features and the criteria. The only questions remaining are how to solve for the predictor and what to predict. To some extent, the proposed solution answers these questions simultaneously.

Let us suppose, for the moment, that we have N observations on the actual workload, y, and simultaneous measurements of n features $x_1, \ldots x_n$ with which we wish to predict the scalar y. A popular, objective, and solvable approach is to seek the set of weighting coefficients ($a_1, \ldots a_n$) such that the predicted ($\hat{y} = a_1 x_1 + \ldots + a_n x_n$) is closest to the actual y in a least-squares sense, i.e., $\epsilon = \overline{(y-\hat{y})^2}$ (where $^-$ indicates average over the N observations) is minimized. From this point on, we will assume that the x's and y's are standardized (zero mean and unity variance). For standardized variables, the solution to this problem is

$$a = \begin{bmatrix} a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{bmatrix} = \overline{xxT}^{-1} \; \overline{xy}$$

where $\overline{xxT}$ is the covariance matrix (same as correlation matrix since the x_i are standardized). If there is some reason to suspect that higher-order terms in x may have predictive value then these are merely added, viz., $x_{n+1} = x_1^2$, $x_{n+2} = x_1 x_2$, etc., and the solution is identical. Thus, we have a tidy technique for finding the best predictor and we need only decide what we wish to predict.

The whole structure of the study was intended to provide this measure of workload in terms of secondary task performance (miss rate and response time). Both of these measures are clearly sensitive to workload, but there seems no way to establish, a priori, the relative goodness of these criteria.

Simultaneous Least-Squares Prediction

Based on some confidence in the physiological features, we propose predictability as this measure of goodness. That is, we will find the criteria coefficients $b_1, b_2, \ldots b_m$ such that the criterion

$$y = b_1 y_1 + b_2 y_2 + \ldots + b_m y_m \tag{1}$$

is best predicted by the n features

$$\hat{y} = a_1 x_1 + a_2 x_2 + \ldots + a_n x_n \tag{2}$$

That is, given the N simultaneous observations on the m criteria and n features, we will find the $m+n$ coefficients such that

$$\epsilon = \overline{(y-\hat{y})^2} = \overline{(b^T y - a^T x)^2} \tag{3}$$

is minimized.

There are, under rather general conditions, m unique solutions (where $m \leq n$) to this problem. The vector of criteria weights is the eigenvector* for the matrix

$$M = \overline{yyT}^{-1} \; \overline{yxT} \; \overline{xxT}^{-1} \; \overline{xyT} \tag{4}$$

and, in particular, the eigenvector b corresponding to the largest eigenvalue λ of M is the best set of criteria weights.

The feature weights are then specified by

$$a = \frac{1}{\sqrt{\lambda}} \; \overline{xx^T}^{-1} \; \overline{xy^T} b \tag{5}$$

---

*A nontrivial b is an eigenvector of M if there is a scalar λ such that $Mb = \lambda b$ and λ is the corresponding eigenvalue of M.

Details of this derivation and summary of steps in the solution are included in Appendix E.

We were somewhat disappointed to learn that our simultaneous least-squares predictor technique has been used by experimental psychologists for some time under the unlikely name Cannonical Correlation. This did have some advantages, however, since they (Cooley, 1965) have worked out an elaborate significance test for the Cannonical Correlation coefficient, $R = \lambda$. Unfortunately, it requires finding all eigenvalues of M, but the chi squared value is then given by

$$\chi^2 = - [N - 1 - 0.5 (p + q + 1)] \ln \Lambda \qquad (6)$$

where

p = number of features

q = number of criteria

$\Lambda = (1 - \lambda_1)(1 - \lambda_2) \ldots (1 - \lambda_m)$, λ_i are the m eigenvalues of M

with pq degrees of freedom.

Validation

We early discovered that with the extremely large number of features available and good individual correlations, high Cannonical Correlations were easily achieved. This is true since Cannonical R is always higher than the best cross-correlation coefficient in $\overline{xy}^T$ and adding more features always increased R, even though the significance may be decreased. However, when the coefficients (a and b) derived from large feature sets in the main study were applied to the corresponding data from the validation study, we found that to have generalized results the number of features had to be kept low (say $p \leq \sqrt{N}$).

Results

Through a combination of classification ordering and multiple correlation ranking, a "best" subset of 10 features was chosen to predict miss rate and response time. The Cannonical Correlation coefficient is .646 and solution for the coefficients is summarized in Table VI.

From Equation (6) we obtain a chi squared value of 57.5 with 20 degrees of freedom and find we can soundly reject the null hypothesis with $p > .995$ (Table E-1).

Application of these weights to the validation data results in a correlation coefficient between y and $\hat{y}$ of .569. To estimate the significance of this result, consider y and $\hat{y}$ as simply n = 20 pairs of points. From Table E-2 the null hypothesis can be rejected with $p > .99$.

TABLE VI. - LEAST-SQUARES PREDICTORS FOR MISS RATE AND RESPONSE TIME

Correlation coefficient = .646

Chi squared = 57.47

Reject null hypothesis with $P > .995$

Measured workload index

Coefficient	Feature	Feature number
b_1 = .780	y_1 = Response time	3
b_2 = .626	y_2 = Miss rate	2

Predicted workload index

Coefficient	Feature	Feature number
a_1 = 1.183	x_1 = Mean respiration amplitude	6
a_2 = -.946	x_2 = Mean respiration interval	12
-.573	VER latency of overall max.	63
.560	VER amplitude P_2	80
-.514	S. D. respiration amplitude	21
.452	S. D. VCG R-T interval	31
-.347	S. D. VCG T-wave amplitude	27
.289	S. D. VCG R-R interval	33
-.266	Respiration rectification	20
-.189	Mean VCG T-wave amplitude	26

From the above results it is concluded that we have evolved a statisticall significant predictor of secondary task performance which has proven genera zable to an entirely different set of data. Before commenting on the physiolo gical features selected, the following questions should be asked of the predic tion system:

1) How do subjective rating and tracking compare with response time and miss rate as criteria?

2) What is the Cannonical Correlation for this feature set in the validation study data?

3) What is the effect of reducing the feature set?

The first of these questions is answered in Table VII. It is reassuring to note that subjective rating, which we presume measures workload, is more heavily weighted than tracking error, which measures performance, and that both of these are more important in this predictor than miss rate or response time. It might be well to recall that the weights are computed on the basis of standardized variables which implies that a change in subjective rating which would change y by .809 (1 σ) would occur with probability equal to a .519 change due to tracking error. Although some changes in feature weights occurred, the change is not overwhelming.

If the zero variance binary and task number are used as criteria, the resulting criteria weights are:

.750 Binary classification

.430 Subjective rating

.423 Task number

-.270 Response time

It should be noted that $R_c = .787$, $\chi^2 = 132$, df = 40, $p > .995$, again without substantial changes in the feature weights.

Using the original 10 features and 2 criteria and applying the system to the validation study, we obtain $R_c = .839$, $\chi^2 = 22.3$, df = 20, $p > .750$ with criteria weights

.988 Response time

-.157 Miss rate

Thus, we observe a higher correlation but lower significance and a large differential criteria weight.

As to the question of feature set reduction, the following results are presented starting with the original feature set and three criteria (Table VIII). Based on the feature weights, successive features were removed from the bottom of the list:

TABLE VII. - LEAST-SQUARES PREDICTORS FOR FOUR CRITERIA

Correlation coefficient = .768

Chi squared = 106.37

Reject null hypothesis with $P > .995$

Measured workload index

Coefficient	Feature	Feature number
.809	Subjective rating	4
+.519	Tracking error	1
-.197	Response time	3
+.195	Miss rate	2

Predicted workload index

Coefficient	Feature	Feature number
-.773	Mean respiration interval	12
-.538	S.D. VCG T-wave amplitude	27
-.504	S.D. VCG R-T interval	31
.441	Mean respiration amplitude	6
.375	VER amplitude P_2	80
-.227	VER latency at overall max.	63
.170	Mean VCG T-wave amplitude	26
.116	S.D. respiration amplitude	21
.043	S.D. VCG R-R interval	33
.028	Respiration rectification	20

TABLE VIII. - LEAST-SQUARES PREDICTORS STARTING SET FOR FEATURE SET REDUCTION

Correlation coefficient = .754

Chi squared = 91.26

Reject null hypothesis with $P > .995$

Measured workload index

Coefficient	Feature	Feature number
.980	Subjective rating	4
.188	Miss rate	2
-.069	Response time	3

Predicted workload index

Coefficient	Feature	Feature number
-.748	Mean respiration interval	12
.538	S.D. VCG R-T interval	31
-.475	S.D. VCG T-wave amplitude	27
.465	Mean respiration amplitude	6
.288	VER amplitude P_2	80
-.240	VER latency of overall max.	63
-.091	S.D. VCG R-R interval	33
.087	Mean VCG T-wave amplitude	26
.018	S.D. respiration amplitude	21

Feature	R_c	χ^2	df	p
All 10	.754	91.3	30	p >>.995
Best 7	.752	73.6	21	p >>.995
Best 5	.732	63.6	15	p >>.995
Best 3	.667	48.5	9	p >>.995
1 feature, 1 criteria	.635	---	---	p >>.99

Workload Predictors

The physiological features which comprise this subset include

- 4 respiration features
- 4 VCG features
- 2 Evoked-response features

Plots of the mean and standard deviation by task number for these 10 are included in Figures 33 through 42. (Similar plots for the criteria variables were presented earlier in the test; see Figures 7, 9, and 11.) The plot of mean amplitude, low (Figure 33) is striking in its similarity to tracking error (Figure 44) and relatively small variance. It also exhibits a "high" value on task 3, which is evident in subjective rating, miss rate, and response time. This feature is the average respiration amplitude after the signal has been low-pass filtered at 0.14 Hz.

Mean interval, high is 1/respiration rate as defined by filtering the signal from 0.125 to 1.3 Hz. The respiration rate is thus positively correlated with workload. Again, the "high" rate on task 3 is evident, and the relationship of tasks 7, 8, and 9 seems more like response time than tracking error.

Standard deviation of rectification pieces is essentially the same as standard deviation of amplitude. The conclusion is clearly that the regularity of respiration decreases with increasing workload. The easy-to-hard dictomy is particularly noticeable in this feature.

Rectification is a measure of total ventilation (rate x amplitude) and exhibits roughly the same behavior as the other respiration features.

For the electrocardiogram features the comparison with the criteria is not nearly so strong. The R-T interval is the time between the left vertical excitation (R-wave) and the repolarization (T-wave). The feature plotted in Figure 37 is a measure of the variance of that interval.

T-wave amplitude is measured from the T-P baseline to eliminate the effect of baseline wandering. Mean T-wave amplitude (Figure 38) shows a

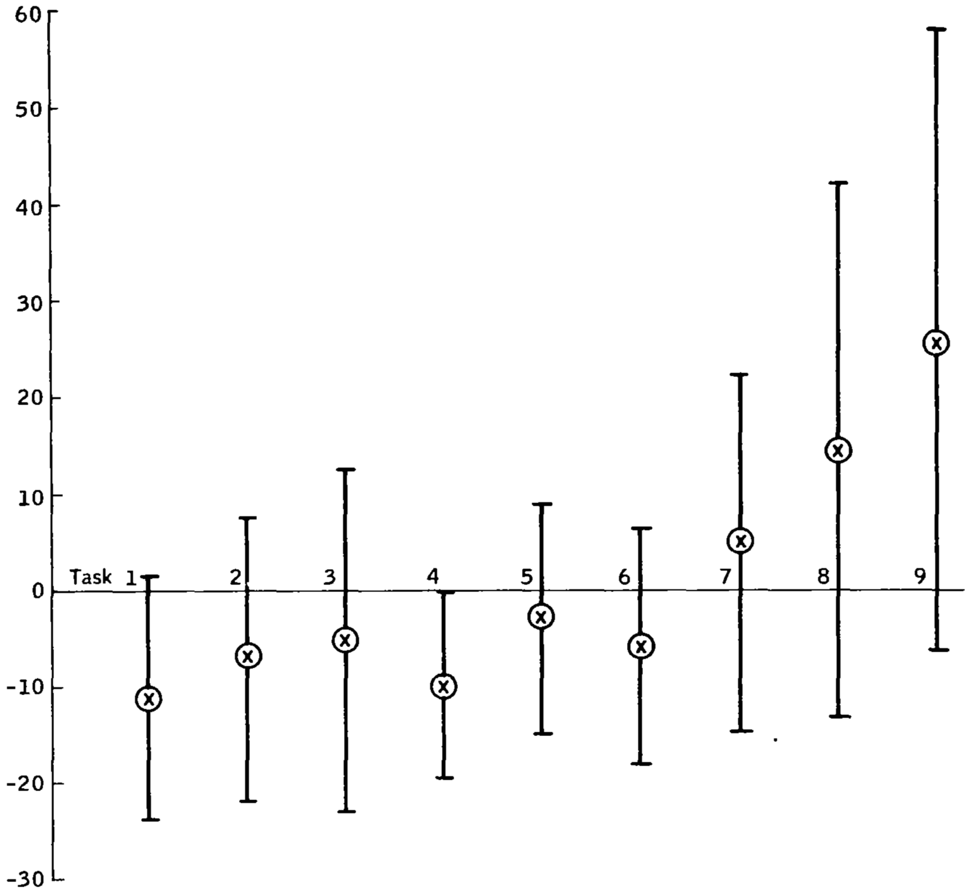


Figure 33. Mean Amplitude, Low

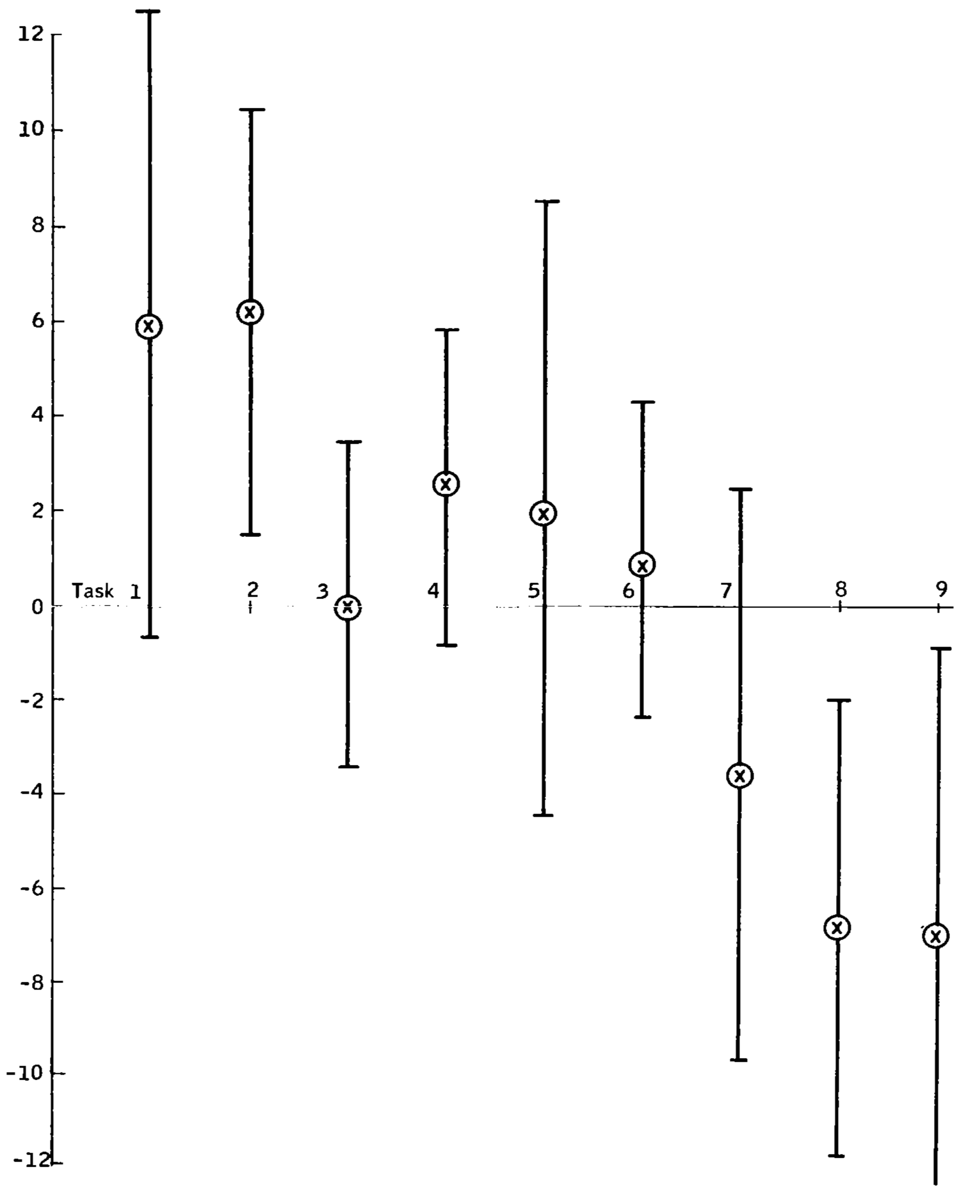


Figure 34. Mean Interval, High

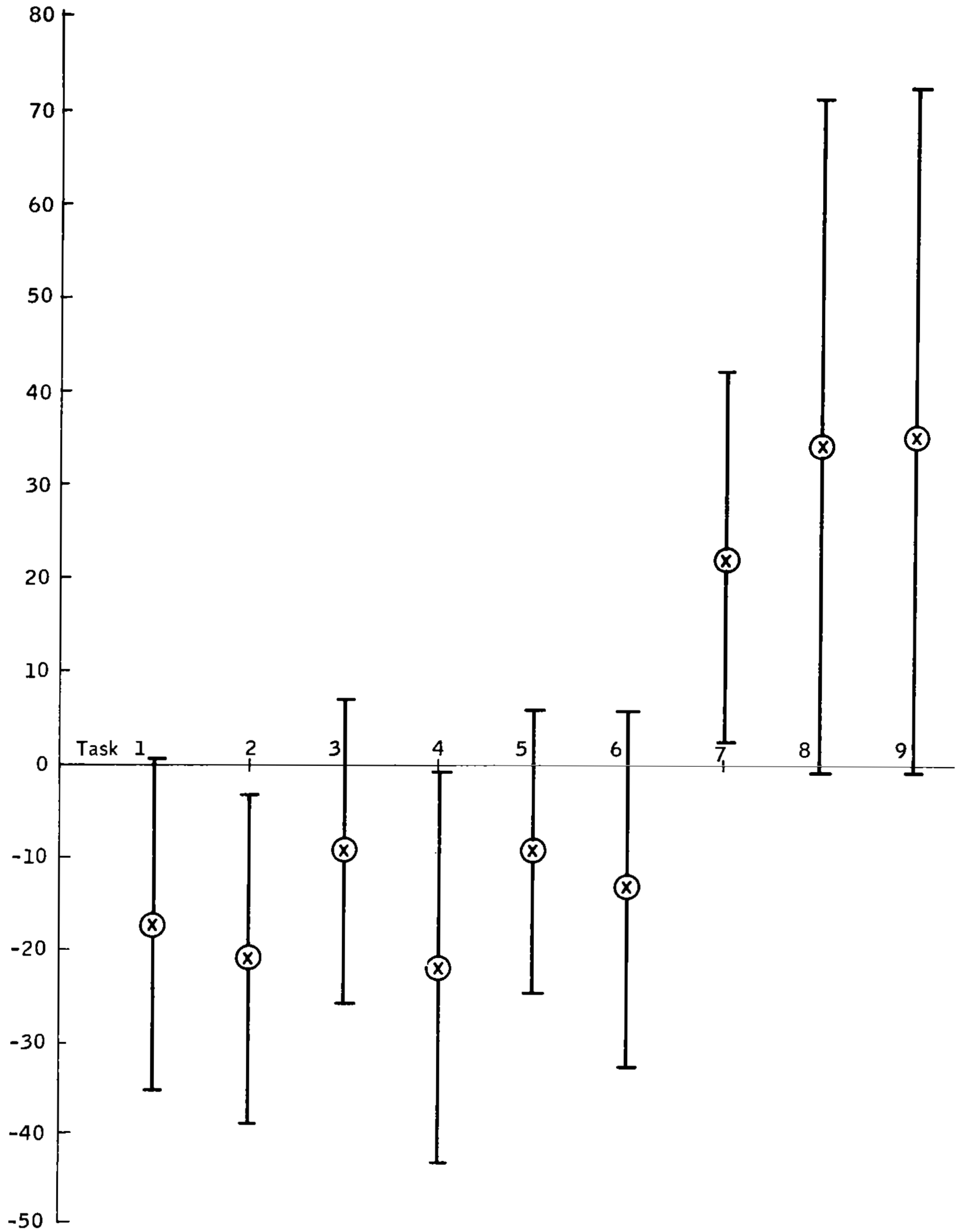


Figure 35. S.D. Rectification Pieces, High

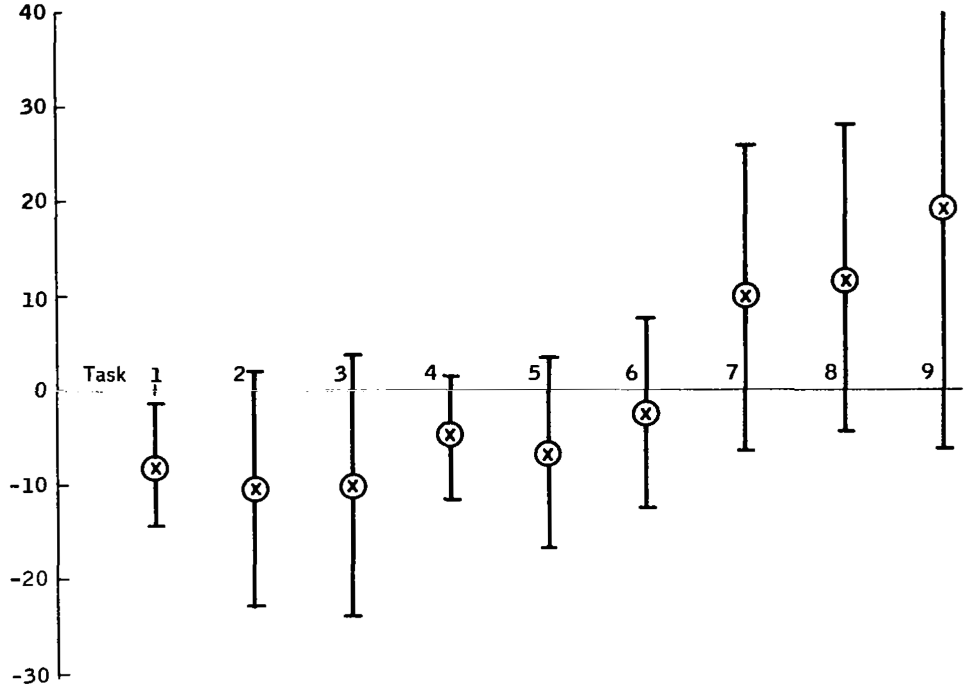


Figure 36. Rectification, High

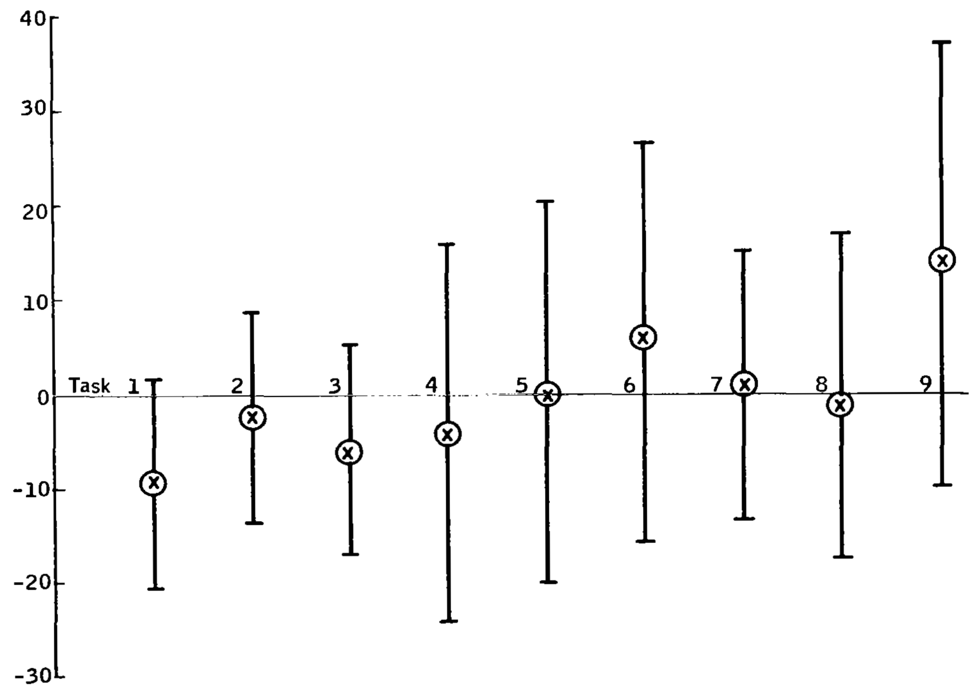


Figure 37. R-T Interval S. D. (Seconds)

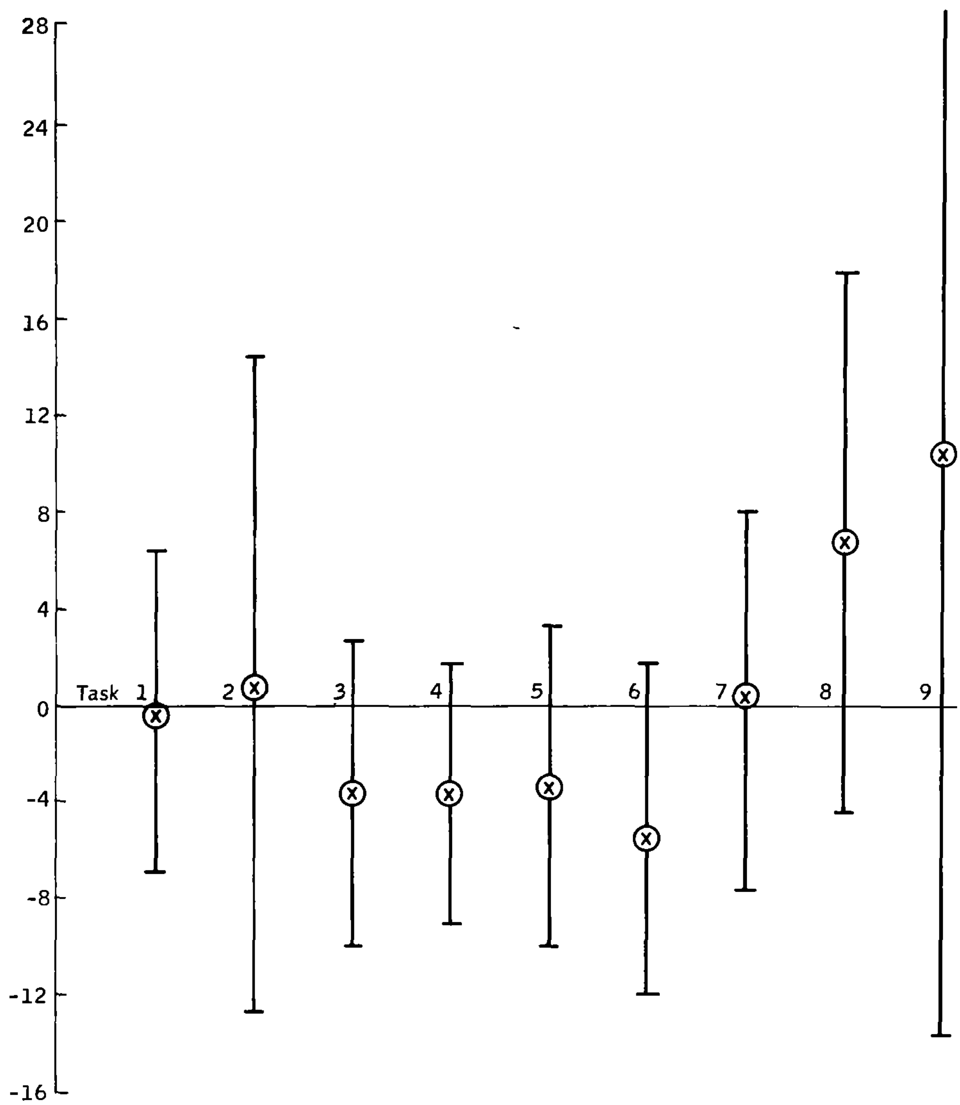


Figure 38. T-Wave Amplitude Mean (Millivolts)

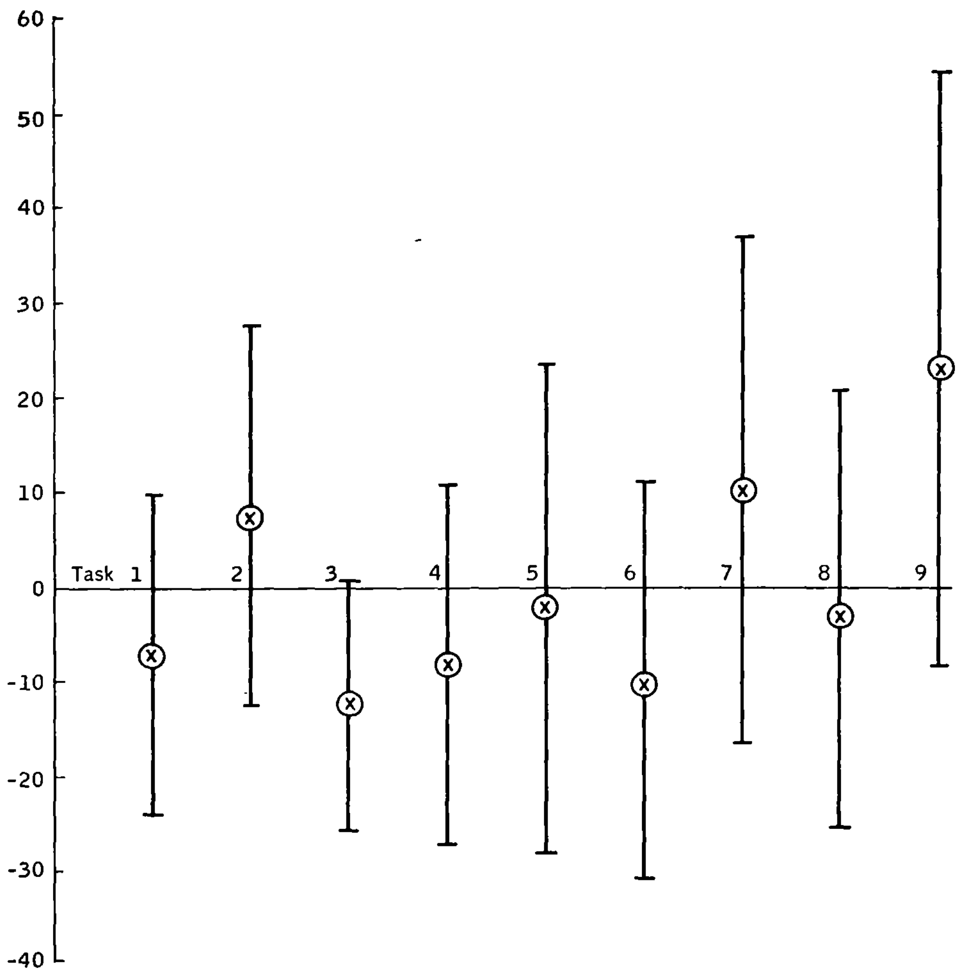


Figure 39. T-Wave Amplitude S.D. (Millivolts)

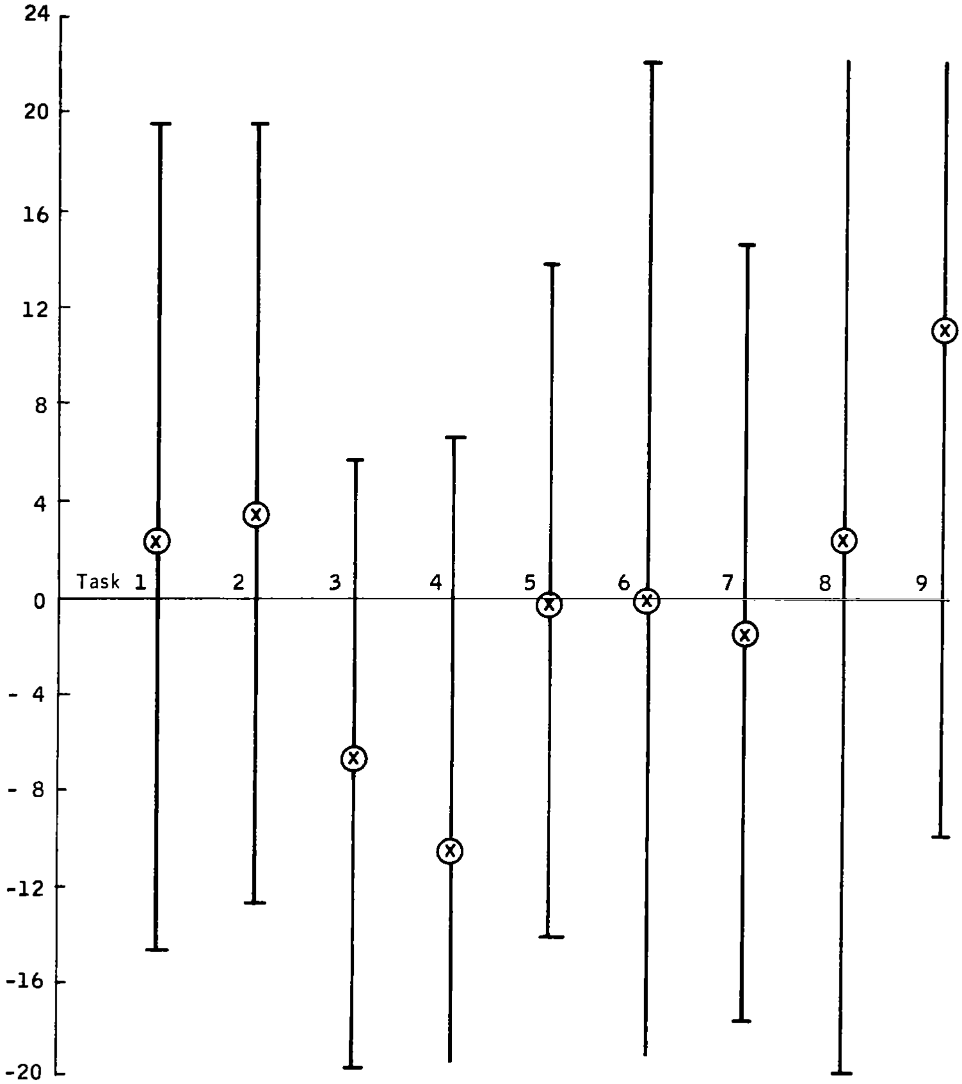


Figure 40. R-R Interval S.D. (Seconds)

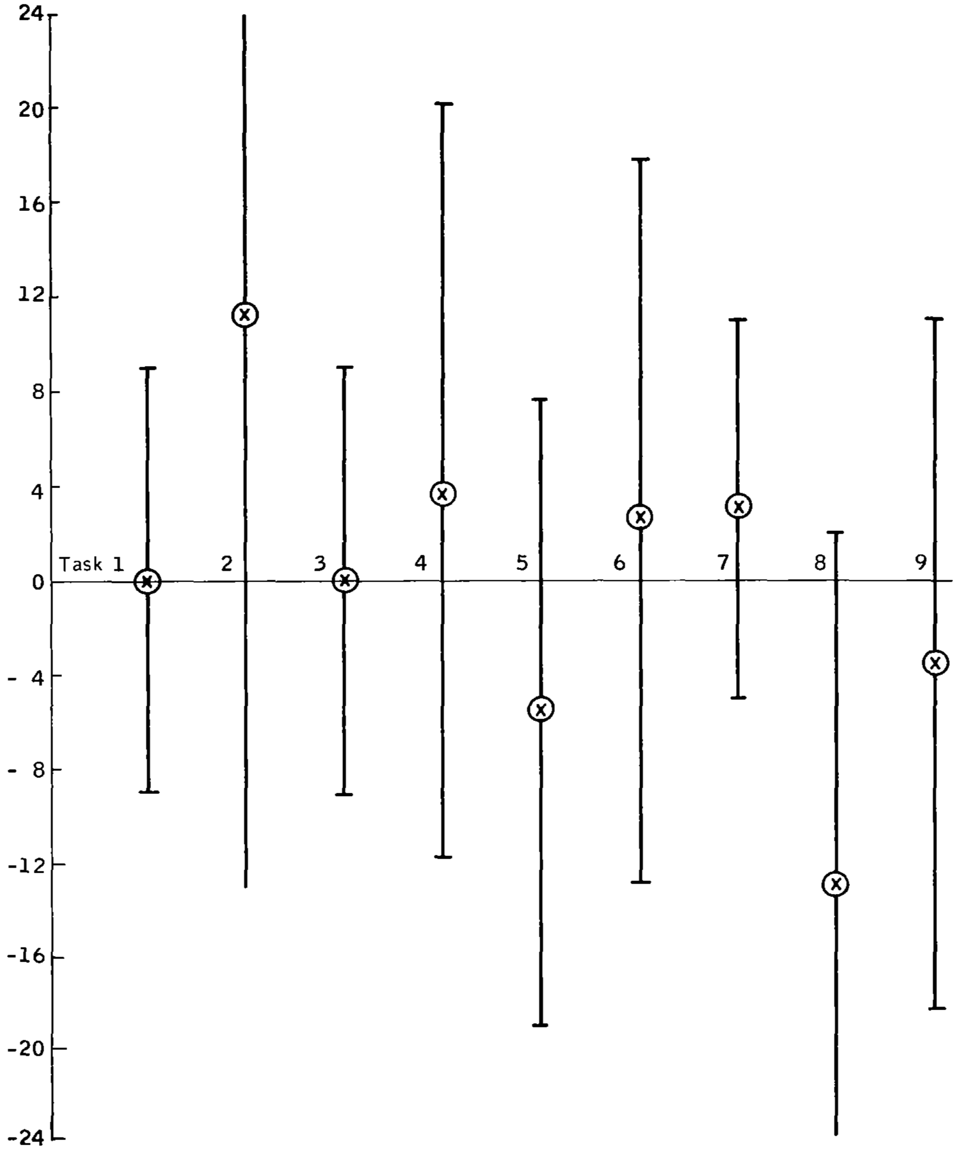


Figure 41. Latency Overall Max. (Milliseconds)

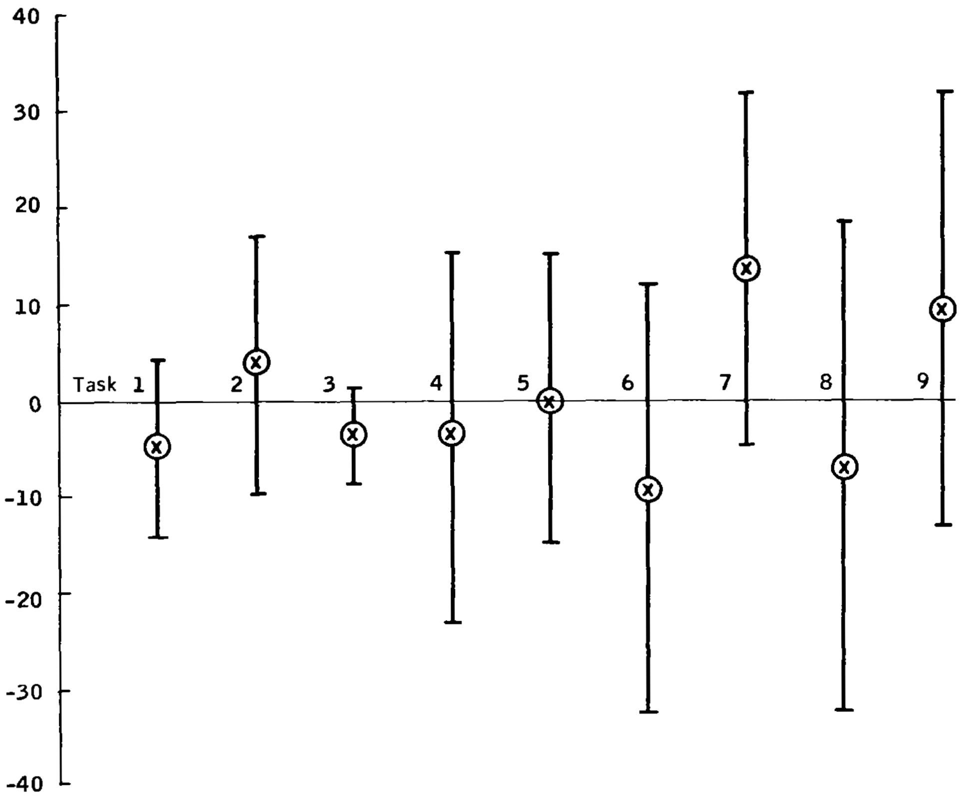


Figure 42. Sequential Max. 3 (Microvolts)

nice tracking error kind of behavior except for tasks 1 and 2. These not withstanding, we can conclude that T-wave amplitude increases with workload.

The variance of T-wave amplitude also seems to roughly increase with workload. For convenience, we measured the mean and variance of the R-R interval rather than rate. The standard deviation of the R-R interval is plotted in Figure 40. The across-subject variance is very large, but again, excluding tasks 1 and 2, an increase in the R-R interval standard deviation is evident.

The visually-evoked response features were generally insignificantly correlated with the criteria variables. Two of the 29 evoked-response features were found to improve the Cannonical Correlation and were given substantial weights (Table VI). The overall maximum was one of the features extracted, and its latency past the stimulus was one of the features selected (Figure 41). The latency decreases with workload, but the intersubject variance is large.

Sequential max. 3 is the maximum which was within or nearest to the interval from 187.5 to 275 ms. It was extracted to correspond to the P_2 (second positive) wave. It shows a positive correlation with workload and a relatively modest intersubject variance.

Summary

We have answered the "What to predict?" question by selecting features and validating using only miss rate and response time and then added the other criteria using the original subset of 10 features.

The "How to predict?" was attacked with least-squares linear prediction which is precisely the Cannonical Correlation solution. The result is a set of weighting coefficients for the features and criteria and an overall correlation coefficient.

The features selected include four respiration features, four from electrocardiogram, and two from evoked response.

CONCLUSIONS

Some of the salient features of this study which we feel represent new or substantially improved techniques include:

1) A simple, sensitive, nonloading secondary task

2) A subjective rating which agrees with other secondary task measure but has less intersubject variance

3) A multichannel physiological monitoring and recording system for respiration vectorcardiogram, electromyogram, electroencephalogram, skin impedance, and subject performance

4) A set of automatic feature extraction software which transforms the analog data base into meaningful features

5) Very good separation results using a pattern recognition system, assuming the data to represent a two-class problem

6) Use of simultaneous least-squares prediction to arrive at a statistically significant, validated workload index and the physiological features which best predict it.

The application of our pattern recognition system to the final data base as a two-class problem resulted in 94-percent separation using the eight best features. The same result was achievable using only respiration and electrocardiogram features. On the validation study data, 100-percent separation was achieved with five features.

Correlation studies showed that most of the respiration features and electromyogram were highly correlated (.5) with primary and secondary task performance. Vectorcardiogram features also showed significant correlations (.2 to .3), but the skin impedance and evoked-response features exhibited low correlations.

A best subset of 10 of the original 84 features was selected and the least-squares linear predictor derived. For 10 features and two criteria the predicted versus observed workload index was correlated with R = .646, significant at the .005 level. The weighting coefficients for standardized variables are:

- Measured Index

  .780 + response time + .626 miss rate

- Predicted Index

  1.183 rspiration amplitude -.946 respiration interval

  -.573 VER latency at max. +.560 VER amplitude of P_2

  -.347 σECG T-wave amplitude + .298 σECG R-R interval

  -.266 total respiration ventilation -.189 ECG T-wave amplitude

Application of these coefficients to the validation study data resulted in a correlation coefficient R = .569, which is significant at the .01 level.

If tracking error and subjective rating are added to the workload index, the weights are

.809 SR + .519 TE - .197 RT + .195 MR

and if tracking error (which is a performance versus workload measure) is deleted, the weights become

$$.980\ SR + .188\ MR - .069\ RT$$

In this last case (10 features and 3 criteria)

$$R = .754$$

Figure 43 presents predicted (physiological) workload versus measured workload for this case by task.

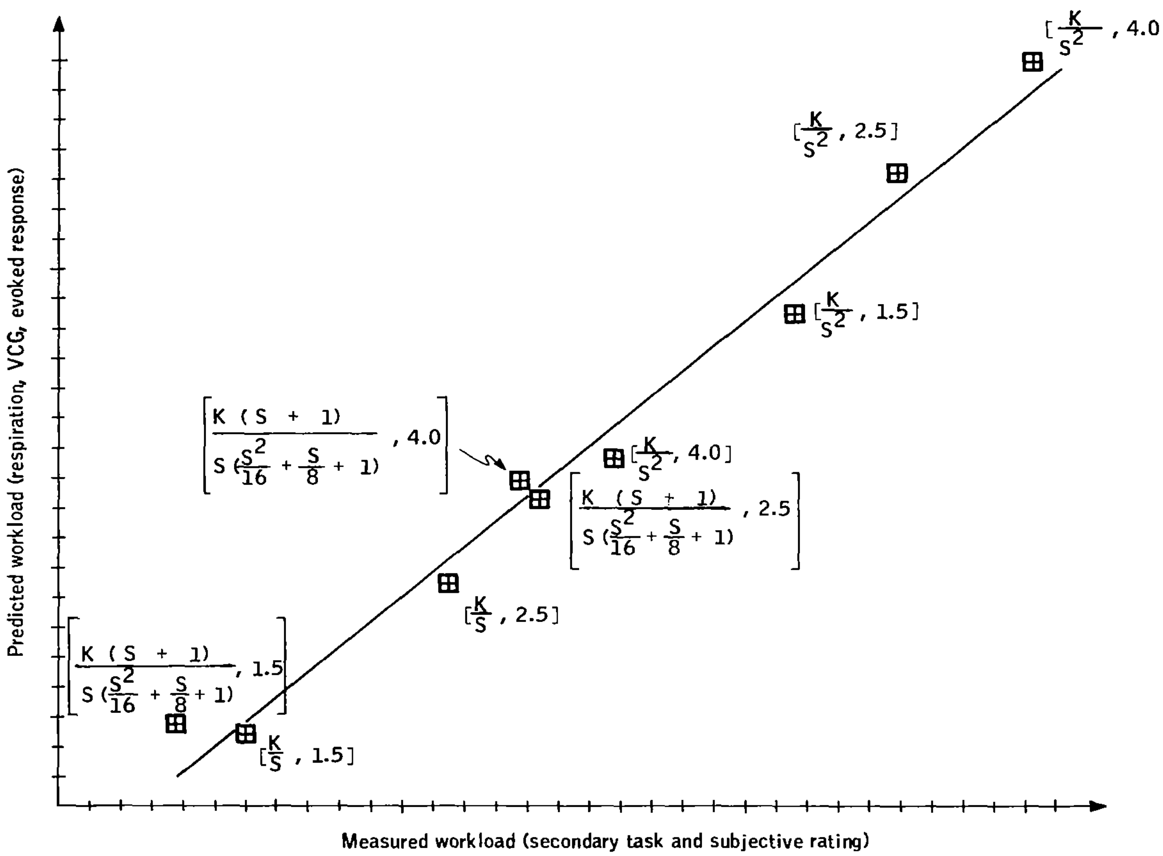


Figure 43. Predicted versus Measured Workload Averaged by Task

Thus, for our experimental situation, the following workload index is recommended:

$$MWI = 1.0 \times \text{subjective rating} + 0.2 \times \text{miss rate} - 0.1 \times \text{response time}$$

The physiological features in the final subset of 10 include four from respiration, four from electrocardiogram, and two from evoked response. Of these respiration is clearly the strongest. Evoked response is subject to very large intersubject variance and its usefulness is limited on that basis. Considerable effort went into a system to measure skin impedance at several frequencies, fit a model, and compute model parameters. Although data fit the circular arc, and hence the model, remarkably well, neither the model parameters nor the magnitudes of impedance showed any significant correlations with workload.

This study has shown that this approach (multichannel monitoring, automatic features extraction, feature selection, and least-squares prediction) represents a viable method for measuring pilot workload. Further, it is concluded that a system which includes only high-quality respiration, scalar electrocardiogram, and electromyogram information can achieve this measurement.

APPENDIX A

MEASURES OF RESERVE CAPACITY

The measures which can be used for establishing the reserve capacity of an operator are shown in Figure A-1. Review articles by Brown (1964) and Knowles (1963) summarize the more important studies relating to these measures.

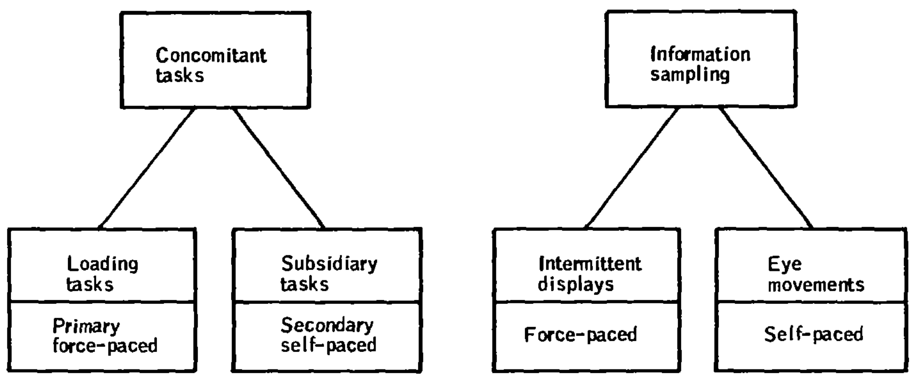


Figure A-1. Methods of Measuring Informational Workload

CONCOMITANT TASKS

Concomitant tasks can be of two types, loading tasks or subsidiary tasks. Loading tasks are characterized by two features. First, by appropriate instruction, the subject is required to perform the loading task at the expense of his performance on the primary task. Second, the loading task is force-paced, i.e., the subject does not control the rate at which he must respond.

The subsidiary task, by instruction, is to be performed by the subject only when he feels he can respond with no decrement in his performance on the primary task. Thus, the subsidiary task is self-paced.

Concomitant tasks can involve the same or different sensory or motor channels used in the primary task, depending on whether the experimenter is concerned with sensor channel capacity, central capacity, or control channel capacity. Determining which sensory/motor channels should be used for concomitant tasks depends on which sensory/motor channels are used for the primary task. These determinations, and the rules for them, form an important area of investigation in developing workload measurement techniques.

APPENDIX A

Subsidiary Tasks

The rationale for the use of the subsidiary tasks is that as the information processing load of the primary task is increased, the operator's information rates on subsidiary tasks are decreased. If it is assumed that these rates are inversely proportional, then a direct measure of primary task workload can be obtained. Ekstrom (1962) used this method in evaluating various control systems for an aircraft using a self-paced, choice-reaction subsidiary task. If, when also performing the primary task, the subsidiary task response was reduced to 50 percent of the level obtained when performing the subsidiary task alone, she concluded that the operator needed only 50 percent of his attention to perform the primary control task. She found that, although measured system performance for two different control systems was the same, one control system required much less operator attention. Knowles and Rose (1963) used a similar choice reaction task to evaluate the perceptual load of two crewmen performing a simulated lunar landing. They found significant differential task loadings between the crewmen, which indicated a need to reallocate crew functions to avoid task overload.

Loading Tasks

The rationale for using loading tasks is that, as the information processing demands for the loading task are increased, performance on the primary task will deteriorate. To determine the reserve capacity of an operator at some specified minimum performance level on the primary task, the demand of the loading task is increased until the primary task is reduced to the selected level of performance. The information rate on the loading task then represents the operator's reserve capacity, since he is performing at this level while maintaining the selected performance level on the primary task. In other words, the capacity used on the loading task could be applied to another ("second primary") task by substituting the "second primary" task for the loading task. In effect, the loading task represents information processing requirements of other primary tasks. Since the loading task is force-paced, the problem of operator response bias can be avoided.

Garvey and Taylor (1959) required subjects to perform such loading tasks as mental addition while tracking with two different control systems. They found that the loading tasks had differential effects on the performance of the two systems. They did not attempt to get a quantitative measure of reserve capacity by systematically varying the information rates in the loading tasks.

Glucksberg (1963) used loading tasks involving information input through either the visual, auditory, or cutaneous sensory modalities. The primary task was to track a visual signal on a rotary pursuit device. The loading tasks involved both simple and choice reactions to the three classes of stimuli. Tracking performance, measured as time on target, was relatively unaffected by loading tasks not involving the visual system. Tracking performance deteriorated, however, with visual loading tasks. This study supports the earlier statement that the reserve capacity measured in one sensory

modality may not be generalized to all sensory modalities. This also illustrates the importance of selecting the sensory/motor modalities of the loading task in relation to those used on the primary task.

INFORMATION SAMPLING MEASURES

The rationale for thus determining information workload is that the measurement of forcing-of-information sampling frequencies and durations permits quantification of the demand or relative performance requirements placed on an operator by various tasks.

One information sampling method used to establish reserve capacity is the intermittent display of information. In actual practice, intermittent sampling of a display is common since the operator must usually divide his attention among several information sources. Intermittent displays are used to determine the time an operator has available to sample information sources other than the primary task source. Stated simply, the critical assumption of the intermittent information technique is that if an operator can perform a defined tracking task at a minimum (but acceptable) level when the display information is available to him only 30 percent of the time, it is assumed he can direct 70 percent of his attention to other information sources. This assumption may be unwarranted. It seems likely that as the percent or time that information is presented decreases, the operator's internal information processing workload actually increases, compensates somewhat for the lack of information presented, and permits maintenance of a high level of performance. This increased internal workload probably takes the form of mental integrations, differentiations, and predictions to compensate for the missing information. These processes may actually be more complex -- and impose higher internal workloads -- than the ones used when the displayed information is present large proportions of the time. At some point, as information availability continues to decrease, this internal processing can no longer compensate for the lack of information, and then observable system performance begins to degrade. It seems almost certain that the "reserve capacity" available at this point will be substantially lower than the difference between the 100-percent-time information presentation level and the percent-time level at this point. Thus, this technique may greatly overestimate the operator's actual reserve capacity in a task. Because of the inherent nature of this measurement technique, it is restricted in application to the determination of operator reserve capacity for the sensors modality involved in the primary task.

Intermittent displays have been investigated in a number of studies. For example, Katz and Spragg (1955) used irregular and sinusoidal target movement in a pursuit tracking task with intermittent display. The display was illuminated with a $1/20^{th}$ of a second flash over a range of 1 flash every 2 seconds to 4 flashes every second. Tracking performance improved with increasing frequency over the entire range. Senders (1955) used a two-dimensional tracking task and found continued improvement with flash rate frequencies as high as 20 per second. However, performance -- even at 20 cycles per second -- was inferior to performance using a continuous display.

APPENDIX A

Since our interest involves the relatively low frequency of voluntary eye movements, high frequencies are not of great concern here. These studies emphasize the point, however, that performance on continuous tracking tasks can be expected to deteriorate if the information source is interrupted.

APPENDIX B

PHYSIOLOGICAL MEASURES OF WORKLOAD

LITERATURE SURVEY

For some time investigators have been searching for a physiological response (or a combination of physiological responses) having a quantitative relationship to some behavior state of the human operator. During the first half of this century, extensive research efforts were devoted to correlating measurements of single responses, such as galvanic skin response, blood pressure, heart rate, skin temperature, etc., with various states. Many of these measures continue to be widely used by physiologists in psychophysiological research, such as classical conditioning, emotional reactivity, and arousal. Results of these investigations suggest that little value can be derived from the low correlations between individual autonomic responses and the level or degree of activation. This is probably attributable to the high intersubject and intrasubject variability of these measures.

It would be convenient if there were a single, easily-measured physiological response having a defined quantitative relationship to information workload. If such a measure were available, it perhaps could be used to measure operator reserve capacity in operational situations. However, the results to date in this field have been generally discouraging.

Fraser (1964), using three experienced RCAF test pilots in low-level, high-speed flights over rough terrain, noted no relationship between heart rate and severity of the flight, as measured in terms of peak acceleration and frequency. However, he reported a marked and sudden increase in the "S-S" interval of the ECG (particularly marked in one subject). This change in the "S-S" interval, according to Fraser, "bore relation" to a severe acceleration occurring 1 or 2 seconds previously. Heart rates, though, were high and varied markedly from sample to sample throughout each flight (range, 88 to 114 per minute). Respiratory rates also tended to be high (up to 30 per minute) throughout each flight. No significant correlation was obtained between respiratory rates and severity of impact and frequency.

On the other hand, Soliday and Schohan (1965) reported that heart rate correlated (Spearman rhos) +.58 with rms "G" and +.53 with rms altitude error. The subjects were eight experienced jet test pilots involved in piloting (primary) and navigational (secondary) tasks while "flying" a TFX-type aircraft in a simulated low-altitude, high-speed mission. They also obtained a correlation of +.73 between respiratory rate and rms "G" and a correlation of +.69 between respiratory rate and rms altitude error ($P = .05$).

Guedry et al. (1964) reported no indication of changes in EKG or blood pressure before, during, or after a dial test which was used as a stressor. This occurred in a two-week rotation run in the Pensacola Slow Rotation Room rotated at 3 rpm. The dial test involved five dials placed so that the subject was required to rotate his head and body through different complex arcs to view the dial and adjust the dial indicator.

APPENDIX B

Psychophysiological measures such as heart rate, skin resistance, integrated electromyogram, respiratory rate, peak inspiratory flow, total lung ventilation, and end tidal CO_2 have been used, either separately or in combination with secondary task performance to estimate primary task performance. Benson, et al (1965) reported that whereas the measures of heart rate, integrated EMG, and pulmonary ventilation each showed a significant increase when the secondary task (acknowledgement of the presence of an intermittent light) was introduced, none of these measures, by themselves, indicated significant difference between the two displays used (counter-only displays, and counter/pointer display). An exception, though, was noted in galvanic skin resistance, where a significant difference between the counter-only and counter/pointer display was noted using the Wilcoxon nonparametric test ($P = .04$). However, this difference was not significant when an analysis of variance was applied.

Benson, et al (1965) found that to demonstrate any difference between tasks, it was advantageous to combine all psychophysiological measures as representing autonomit and somatic nervous system activity and analyzing mean task differences. Only by analyzing these combined measures were the experimenters able to rank order the tasks in terms of operator demand.

Lacey introduced this approach in the assessment of psychophysiological variables in 1950. He suggested that specific emotions could very well be correlated with patterns of autonomic responses, and that their relationship could best be expressed through response profiles among several autonomic measures. Lacey stated that patterning of autonomic reactions is a variable possibly more important than average reactivity itself. Using T-scores and regression models, Lacey demonstrated that response patterning occurred between several psychophysiological measures. Lacey (1963) distinguished two classes of visceral-autonomic variables. The first class deals with the "organism's responsivity" dimension, which stems from measures of the variability of steady-state autonomic activity along a "stabile-labile" dimension, and the second class of variables indicates response patterns of visceral-autonomic function.

Pribram (1967) states that cerebral activation is a "change in the state of organization of neural patterns related to the configurational incongruity between input and established neural activity." Behavior arousal is not necessarily expressed as a difference in the amount of neural activity but rather as a temporary "state of disequilibrium," a disturbance of patterns of organism-environment interactions which may result in a different state of organization or disorganization. Changes in the autonomic responses indicate that a reaction to "incongruous" input has taken place, but this does not always reflect the organization of the emotional process.

The literature clearly indicates, then, the futility of single variable research in this area. We consider the relationship of specific psychophysiological responses and workload a complex multivariate problem that can best be assessed by applying modern computer techniques, and using signal classification and pattern recognition analysis of numerous autonomic measures.

APPENDIX B

CARDIAC CONTROL

The fundamental role of the circulatory system is to supply blood to the capillaries, thereby permitting exchange of metabolites with the tissue cells.

The heart is a double pump with the main chambers (ventricles) contracing almost simultaneously. The blood is not merely pushed out; it is virtually wrung out of them by the squeeze (systole) of the spirally-arranged cardiac muscle.

The left ventricle, which carries most of the circulatory load since it pumps against five times as much pressure as the right ventricle, forces blood through the aortic valve into the aorta. The elasticity (or inversely, compliance) of this artery provides storage for the blood, as well as for the potential energy, so that at the end of the heart's filling phase (diastole) the aortic pressure is still about 80 mm Hg.

In the resting adult, the average heart rate is 70 beats per minute and stroke volume 70 ml so that the cardiac output (CO) is about 5 liters per minute. In times of stress (e. g., maximal exercise) CO may reach 25 liters per minute or as high as 35 liters per minute in a trained athlete. In either case. the maximum heart rate is approximately 180 beats per minute, so the stroke volume must be on the order of 150 to 200 ml.

The cardiovascular system can adjust impressively to stress. For example, the skin, which normally receives 0.2 to 0.3 liters per minute may receive 5 to 7 liters per minute during severe heart load. Let us briefly examine some of the mechanisms of cardiac control.

Intrinsic Control Mechanisms

For an isolated myocardium (no neurological or hormonal control) the energy of contraction exhibits accommodation to end diastolic volume (heterometric) and to sustained load changes (homeometric). The former is the well-known Frank-Starling mechanism. From simple mechanical considerations, the wall tension to produce a given fluid pressure in the ventricle varies with the square of the radius. Starling observed, however, that the energy of cardiac contraction was proportional to the initial length of the muscle fibers. Thus, the regulatory mechanism which maintains the balance between right and left cardiac output is designed into the lowest level of the system.

In addition to the heterometric regulation which operates with each contraction, there is a slower responding autoregulator mechanism which tends to return the operating point to the nominal end diastolic volume (EDV). Since this accommodation tends to keep EDV constant, it is termed homemetric autoregulation.

Heterometric regulation can be represented by a cardiac function curve (stroke work versus EDV); the homeometric correction amounts to a shift in the function curve (Figure B-1).

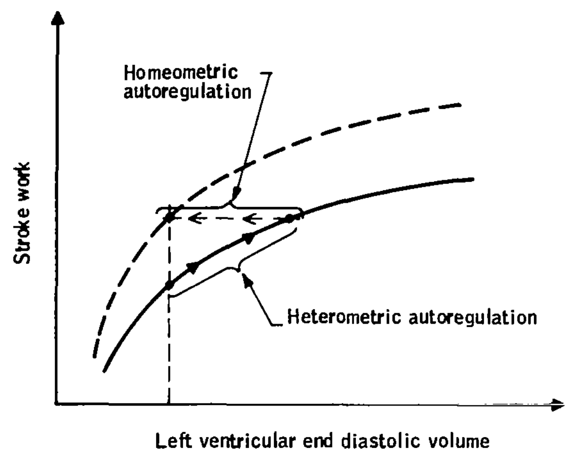


Figure B-1. Cardiac Function Curve and a Shifted Curve (---) Showing Intrinsic Autoregulatory Response to a Change in Load

This adjustment may occur in response to changes in aortic pressure, heart rate, and venous return. The advantage of such regulation is that it tends to conserve heterometric regulation and keep small the ratio of systolic to diastolic period.

Extrinsic Control

The outside influences on cardiac function include neurological, hormonal, and fluid mechanical. The autonomic nervous system differs from the voluntary motor system in that it supplies smooth muscle, cardiac muscle, and certain glands - structures over which we ordinarily exercise no control. The autonomic system is separated into two anatomically, functionally, and pharamacologically distinct (yet coordinated) systems: sympathetic and parasmpathetic.

Functionally, the sympathetic system is primarily an emergency system which prepares the body for "fight or flight" in the face of danger. For a fixed heart rate, sympathetic stimulation of cardiac muscle results in increased CO, increased arterial pressure, and reduced end diastolic pressure (EDP). This is achieved in some degree by an increased synchronicity of contraction of the ventricular muscle fibers.

The parasympathetic system is primarily a homeostatic system which tends to promote orderly bodily processes. Parasympathetic stimulation of

the heart - via the vagus nerve - results in decreased stroke work. In contrast to sympathetic stimulation, the vagus largely effects atrial contraction.

The relative effects of sympathetic and vagal stimulation on contractility (in this case referring to peak ventricular pressure) have been empirically determined by Martin and Levy (1967) for a particular experimental situation:

$$\begin{aligned} \text{Contractility} &= 0.109 + 0.165S - 0.09V \\ &\quad -0.057S^2 - 0.12V^2 - 0.035V \end{aligned}$$

where

S = log of sympathetic stimulation

V = log of vagal stimulation

Heart rate is also mediated by vagal and sympathetic stimulation, with the former acting to reduce heart rate and the latter acting to increase it. This relation was quantitatively described by Warner (1967):

$$\text{Rate} = R_v + (R_s - R_v)\,\frac{R_v - R_{min}}{R_o - R_{min}}$$

where

R_v = rate due to vagal stimulation alone

R_s = rate due to sympathetic stimulation along

R_o = rate with zero stimulation

R_{min} = minimum heart rate (usually about 30 beats per minute) achievable with vagal stimulation (further stimulation will stop heart)

Various hormones also exercise control over the rate and contractility of the heart. Most notable are epinepherine and acetylcholine which have effects similar to the sympathetic and vagus nerves, respectively. Epinepherine is produced primarily by the adrenal medulla and affects the heart in remarkably low concentrations (1 part in 10^9).

Neural and hormonal stimulation are thus the control inputs to the heart. They are largely derived from mechanical and chemical transducers located in the heart and arteries, thus providing feedback control of cardiac function.

Under normal resting conditions the most important reflex control signal is that of the mechanoreceptors located in the carotid sinus and aortic arch. These transducers measure vessel wall stretch and thus pressure; increased stretch → increased vagal firing rate → decreased heart rate.

During periods of unusual stress which result in acidosis or anoxia, the chemoreceptors may assume a dominant role in both cardiac and vascular regulation.

In addition to this feedback regulation, there is control exerted by higher centers. Emotions (the "fright-fight-flight" reactions) are mainly apparent in the sympathetic nervous stimulation.

Vascular Regulation

The major seat of resistance in the systemic circuit is the arterioles. Their total cross-section area is approximately that of the aorta, but in laminar flow, resistance is much greater. All vessels except capillaries have a smooth muscle component (vasoconstrictor) which is under sympathetic control. The capillaries can, however, actively modify their own caliber in response to local nervous, hormonal and other chemical and physical stimuli.

Venules and veins contain about two-thirds of the body's 5 liters of blood. Vasomotor constriction in the small veins (venules) thus drastically affects the storage volume and hence the return of the blood to the heart, although it does not appreciably affect the overall resistance. Conversely, the arterioles mediate resistance with negligible effect on the system capacity.

In addition to local resistance changes, there are substantial anatomical and/or physiological shunts for the major organs. In some instances their value is clear (diverting blood from the viscera during exercise stress), but in others it is not (50 percent physiological shunt of pulmonary capillaries during hypervolemia).

Figure B-2 summarizes the functional relationships involved in cardiovascular regulation.

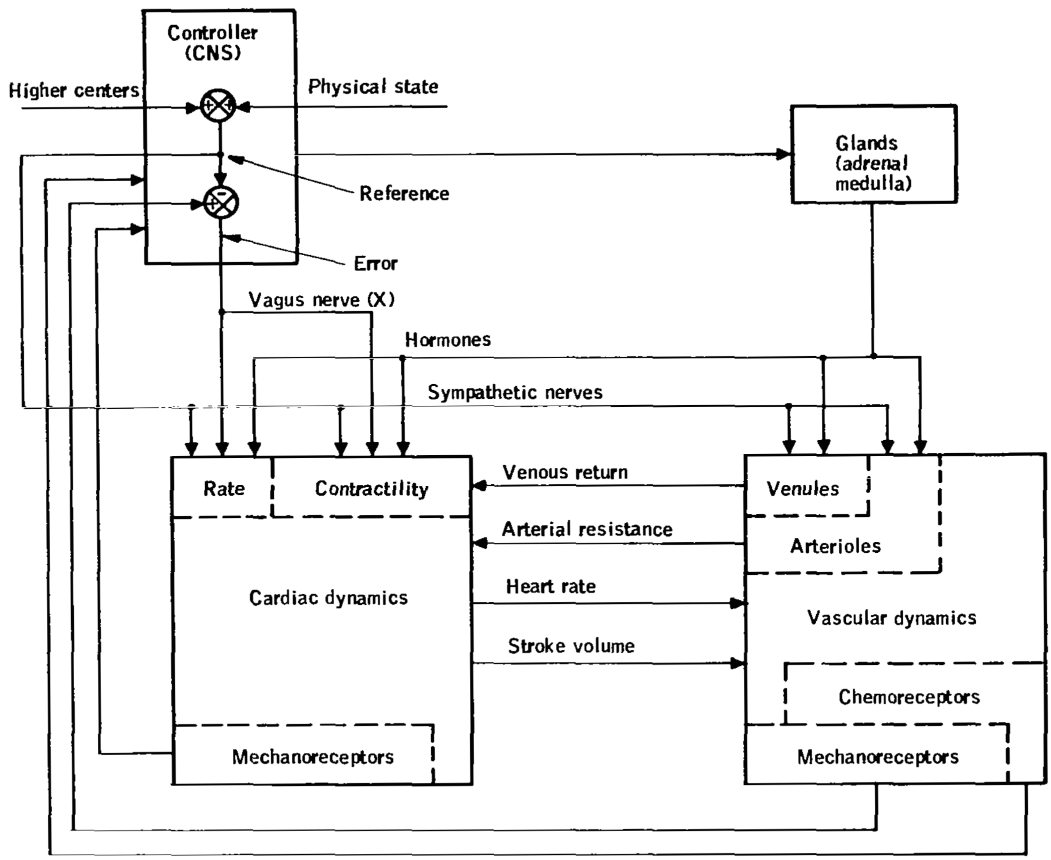


Figure B-2. Summary of Cardiovascular Controls

APPENDIX C

EXPERIMENTAL DESIGN AND EQUIPMENT

TABLE C-1.- FACTORIAL DESIGN - MAIN EXPERIMENT

Subject	Section 1									Section 2									Section 3								
	Run Number									Run Number									Run Number								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	11	22	33	21	32	13	31	12	23	21	32	13	31	12	23	11	22	33	31	12	23	11	22	33	21	32	13
2	21	32	13	31	12	23	11	22	33	31	12	23	11	22	33	21	32	13	11	22	33	21	32	13	31	12	23
3	31	12	23	11	22	33	21	32	13	11	22	33	21	32	13	31	12	23	21	32	13	31	12	23	11	22	33
4	22	33	11	32	13	21	12	23	31	32	13	21	11	23	31	22	33	11	12	23	31	22	33	11	32	13	21
5	32	13	21	12	23	31	22	33	11	12	23	31	22	33	11	32	13	21	22	33	11	32	13	21	12	23	31
6	12	23	31	22	33	11	32	13	21	22	33	11	32	13	21	12	23	31	32	13	21	12	23	31	22	33	11
7	33	11	22	13	21	32	23	31	12	13	21	32	23	31	12	33	11	22	23	31	12	33	11	22	13	21	32
8	13	21	32	23	31	12	33	11	22	23	31	12	33	11	22	13	21	32	33	11	22	13	21	32	23	31	12
9	23	31	12	33	11	22	13	21	32	23	11	22	13	21	32	23	31	12	13	21	32	23	31	12	33	11	22

Note; Numbers in cells refer to treatment combinations as shown below:

Factor	$\frac{K}{S}$	$\frac{K(S+1)\ 16}{S(S^2+85+16)}$	$\frac{K}{S^2}$
1.5	11	12	13
2.5	21	22	23
4.0	31	32	33

TABLE C-2. - RANDOM TASK PRESENTATION - VALIDATION STUDY

Subject	Run Number					
	1	2	3	4	5	6
1	21	12	12	21	21	12
2	21	12	21	12	12	21
3	12	21	21	12	12	21
4	12	12	21	12	21	21
5	12	21	12	12	21	21
6	12	12	21	21	21	12

Note: Numbers in cells refer to treatments:

12 = pitch dynamics $\frac{K}{S}$ switched to $\frac{K}{S^2}$

21 = pitch dynamics $\frac{K}{S^2}$ switched to $\frac{K}{S}$

APPENDIX C

TABLE C-3. - SPECIFICATIONS FOR HONEYWELL BIOMEDICAL AMPLIFIER

Parameter	Amplifier 9	Amplifier 10
Frequency response (3 dB)	0.135 Hz 2040 Hz	0.135 Hz 2090 Hz
Input impedance at 100 Hz	INV = 45 MΩ NINV = 50 MΩ	INV - 36 MΩ NINV - 43 MΩ
Output impedance at 100 Hz	314 Ω	317 Ω
Voltage gain at 100 Hz	1×10^3 min. 40×10^3 max. Variable as per gain chart	Same
Dynamic range at 100 Hz	0.25 mV max. gain 6.9 mV min. gain	0.25 mV max. gain 7.2 mV min. gain
AC output level at 100 Hz	10 V rms max. gain 6.9 V rms min. gain	10 V rms max. gain 7.2 V rms min. gain
Equivalent input noise (broadband)	Shorted input ENV = 2.25 μV	Shorted input ENV = 2.4 μV
Power requirements (charge batteries for 15 hours)	±15 Vdc 35 hours continuous operation on batteries	Same
Dimensions (not including handle or controls)	H = 5.50 in. W = 8.25 in. D = 5.75 in.	Same
Common mode rejection at 100 Hz: $K_v = 10^3$, $E_i = 0.5$ mV	52 dB	58 dB

APPENDIX C

SPECIFICATIONS FOR LOW-FREQUENCY GAUSSIAN NOISE GENERATOR 44.200

- Output load:

  The Model 44.200 is designed to work into a standard analog computer amplifier using a 1 MΩ input resistance.

- Amplitude probability distribution:

  Gaussian (normal) to less than ± (Figure C-1)

- Output spectrum:

  Uniform to ± 0.1 dB from 0 to 35 Hz. Output falls off rapidly above 40 Hz.

- Maximum output level:

  15 V rms (may be decreased by means of built-in attenuator)

- Maximum spectral density:

  Approximately 4 $(V)^2$ per Hz

- D-C unbalance:

  Less than 40 mV with 95 percent certainty

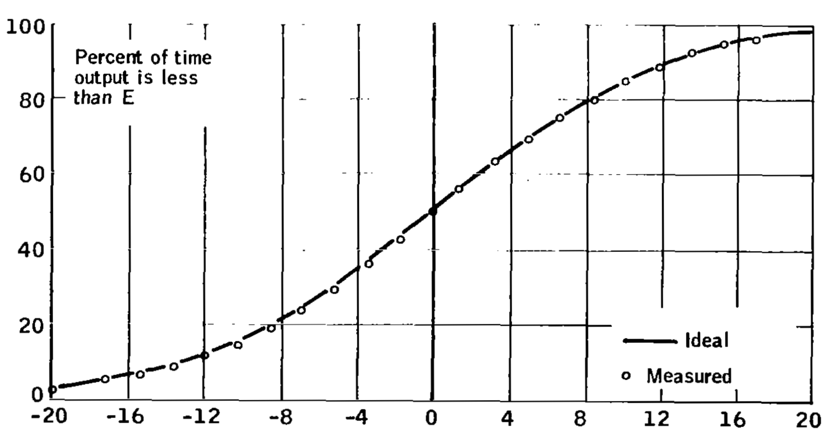


Figure C-1. Cumulative Probability Distribution of Noise Generator Output Showing Agreement Between Measured and Theoretical (Gaussian) Values

APPENDIX C

PILOT WORKLOAD SUBJECTIVE EVALUATION*

Instructions to Pilots

Played at the beginning of each session: "This is the pilot workload assessment project. You have two tasks to perform. The tracking task, which would be your major concern at all times, consists of keeping the horizon line within the indicated limits.

Your secondary task is identifying, and responding to, the two lights above the display. To be scored as correct, your response must be made while the light is on.

Let me again emphasize that your primary concern should be to do the best you can on the tracking. Respond to the lights if, and only if, you feel you can do so without sacrificing your tracking performance."

Questionnaire

The following questions refer to the tracking task. Give only one answer per question.

1. In my opinion the response characteristics of the simulated aircraft were:

- ☐ Excellent, pure, no accidental excitation (0.5)
- ☐ Good, relatively pure (3.5)
- ☐ Fair, somewhat impure (5.5)
- ☐ Quite sensitive, sluggish or uncomfortable (6.5)
- ☐ Extremely sensitive, sluggish or uncomfortable (7.5)
- ☐ Nearly uncontrollable (9.0)
- ☐ Uncontrollable (10.0)

---

* For subject description, see Table C-4 at end of Appendix.

APPENDIX C

II. In my opinion control over the simulated aircraft was:

- [ ] Extremely easy to control with excellent precision (0.5)
- [ ] Very easy to control with good precision (2.5)
- [ ] Easy to control with fair precision (4.5)
- [ ] Controllable with somewhat inadequate precision (6.5)
- [ ] Controllable, but only very unprecisely (7.5)
- [ ] Difficult to control (8.0)
- [ ] Very difficult to control (8.5)
- [ ] Nearly uncontrollable (9.0)
- [ ] Uncontrollable (10.0)

III. In my opinion the demands placed on me as the pilot were:

- [ ] Completely undemanding, very relaxed and comfortable (2.5)
- [ ] Largely undemanding, relaxed (3.5)
- [ ] Mildly demanding of pilot attention, skill, or effort. (5.5)
- [ ] Demanding of pilot attention skill or effort (6.5)
- [ ] Very demanding of pilot attention, skill, or effort (7.5)
- [ ] Completely demanding of pilot attention, skill, or effort (8.5)
- [ ] Nearly uncontrollable (9.0)
- [ ] Uncontrollable (10.0)

APPENDIX C

IV. In my opinion the deficiencies in the simulated aircraft were:

- ☐ Effects of deficiencies on performance are easily compensated for by the pilot (5.5)
- ☐ Moderately objectionable deficiencies (6.5)
- ☐ Major, very objectionable deficiencies (7.5)
- ☐ Nearly uncontrollable (9.0)
- ☐ Uncontrollable (10.0)

V. In my opinion turning off the lights interferred with my performance on the tracking task:

- ☐ Not at all, no interference (0.5)
- ☐ To a negligible extent, did not interfere with tracking (1.0)
- ☐ Some interference, resulted in a few tracking errors (4.0)
- ☐ Moderate interference, caused some tracking errors (5.0)
- ☐ Definite interference, considerable tracking errors resulted (8.0)
- ☐ Nearly complete interference, track was severely impaired (9.0)
- ☐ Complete interference, could not perform on tracking task (10.0)

VI. In my opinion I was able to respond to the lights:

- [ ] Always responded immediately (0.5)
- [ ] Always responded while the light was on (1.0)
- [ ] Always responded, but occasionally too late (2.0)
- [ ] Always responded, but often too late (4.0)
- [ ] Usually responded, and responses were never late (6.0)
- [ ] Usually responded, but responses were sometimes late (7.0)
- [ ] Often failed to respond, but responses were usually in time (8.0)
- [ ] Often failed to respond, but responses were usually late (9.0)
- [ ] Only rarely (10.0)

TABLE C-4. SUBJECT DESCRIPTION

Subject	Age	Years held license	Average fly time work/month	Motivation	Comments
A.H.	33	2	10	Excellent	Extensive experience in flight simulation studies
B.S.	35	14	Presently active	Excellent	Extensive military flying experience. Found subject task fatiguing.
M.G.	30	1	10	Excellent	Well-trained subject
T.L.	28	2	4	Excellent	Well-trained subject
R.T.	33	---	---	Excellent	Learned controls rapidly
T.C.		---	---	Excellent	Subject operated control stick with nondominant hand
E.R.	42	4	8	Excellent	Had trouble with control reversals
M.S.	38	4	5	Excellent	Appeared to perspire more than normal
G.Y.	31	1/2	10	Excellent	Learned controls rapidly
D.B.	32	1/4	15	Excellent	Tended to ignore secondary task completely for difficult primary task condition

APPENDIX D

EXPERIMENTAL RESULTS

TABLE D-1. - MEAN AND STANDARD DEVIATION TRACKING ERROR, DISCRIMINATION PERFORMANCE, AND SUBJECTIVE EVALUATION

Task number	Tracking error	Percent error	Response time	Subjective evaluation
Main experiment (N = 27)				
1	16.88 28.02	5.93 3.45	542.07 40.23	18.60 7.01
2	15.28 16.84	6.95 4.21	542.70 39.48	19.64 5.98
3	19.19 25.87	9.99 8.35	554.77 47.98	27.71 5.63
4	24.22 31.89	10.68 10.36	551.32 50.95	27.08 7.99
5	25.78 28.60	12.42 19.48	560.64 72.09	29.26 8.28
6	40.07 54.64	11.16 9.78	556.89 48.18	31.83 7.69
7	59.65 63.53	23.12 27.88	558.02 76.23	36.78 9.90
8	76.19 56.26	22.26 28.02	604.42 85.92	40.34 8.17
9	85.11 40.95	22.34 27.30	602.19 81.95	41.54 5.98
Validation study (N = 15)				
11	20.15 10.38	7.16 3.53	548.22 51.28	
12	27.84 13.67	9.31 8.35	599.83 72.14	
13	92.96 40.73	22.80 27.87	599.99 93.94	
14	97.67 41.09	21.72 24.77	648.13 93.82	

TABLE D-2. - EMG AND RESPIRATION FEATURE AVERAGES BY TASK (UNNORMALIZED DATA, N = 27)

Feature number	Feature description	Prebase	Task number								
			1	2	3	4	5	6	7	8	9
5	Integrated electromyogram	37.0	39.0	36.0	34.6	43.0	44.6	43.4	43.4	47.6	41.4
	Respiration features										
6	Mean amplitude, low	.204	.144	.130	.138	.137	.157	.156	.181	.186	.197
7	S.D. amplitude, low	.133	.083	.072	.085	.076	.094	.100	.123	.136	.136
8	Mean amplitude, high	.99	1.10	1.00	.97	1.08	1.08	1.13	1.24	1.19	1.23
9	S.D. amplitude, high	.521	.400	.355	.399	.374	.437	.432	.609	.642	.640
10	Mean interval, low	4.15	3.22	3.25	3.37	3.22	3.32	3.27	3.19	3.26	3.38
11	S.D. interval, low	.957	.595	.559	.662	.578	.704	.660	.763	.867	.889
12	Mean interval, high	2.80	2.77	2.76	2.60	2.68	2.68	2.64	2.51	2.41	2.40
13	S.D. interval, high	.899	.614	.547	.650	.572	.651	.582	.599	.693	.711
14	Signal average, low	.011	-.012	-.01	-.01	-.01	-.01	-.01	-.00	-.01	-.01
15	Signal power, low	.109	.080	.072	.077	.075	.088	.092	.107	.133	.11
16	Signal average, high	-.025	-.024	-.025	-.025	-0.023	-.024	-.027	-.023	-.024	-.026
17	Signal power, high	.451	.452	.415	.424	.453	.465	.480	.559	.565	.575
18	Rectification, low	.110	.102	.091	.093	.097	.104	.108	.126	.127	.132
19	S.D. rectification pieces	.266	.166	.145	.170	.153	.188	.200	.245	.272	.271
20	Rectification, high	.762	.932	.832	.842	.926	.923	.990	1.140	1.139	1.172
21	S.D. rectification pieces, high	1.03	.80	.71	.80	.74	.98	.86	1.21	1.28	1.27

TABLE D-3. - VECTORCARDIOGRAM FEATURE AVERAGES BY TASK (UNNORMALIZED DATA, N = 27)

Feature number	Feature description	Prebase	Task number								
			1	2	3	4	5	6	7	8	9
	(millivolts)										
22	R-wave amplitude, mean	.934	.948	.933	.926	.912	.898	.901	.893	.948	.943
23	R-wave amplitude, σ	.134	.127	.122	.113	.117	.122	.114	.121	.115	.125
24	S-T amplitude, mean	.594	.580	.609	.559	.571	.562	.558	.574	.644	.672
25	S-T amplitude, σ	.150	.097	.113	.092	.100	.106	.099	.103	.107	.143
26	T-wave amplitude, mean	.365	.321	.351	.317	.320	.322	.305	.327	.368	.374
27	T-wave amplitude, σ	.097	.061	.077	.060	.066	.068	.061	.069	.069	.093
28	Baseline, mean	-.110	-.089	-.088	-.090	-.082	-.084	-.079	-.035	-.086	-.084
29	Baseline, σ	.055	.025	.034	.028	.033	.030	.027	.033	.031	.033
	(seconds)										
30	R-T interval, mean	.260	.247	.266	.260	.259	.267	.261	.252	.256	.259
31	R-T interval, σ	.044	.040	.048	.044	.046	.049	.050	.046	.047	.052
32	R-R interval, mean	.890	.819	.811	.806	.801	.819	.800	.799	.808	.794
33	R-R interval, σ	.212	.114	.117	.107	.104	.116	.122	.119	.123	.128

TABLE D-4. - SKIN IMPEDANCE FEATURE AVERAGES BY TASK (UNNORMALIZED DATA, N = 27)

Feature number	Feature description	Prebase	Task number								
			1	2	3	4	5	6	7	8	9
	Model parameters										
34	Series resistance (kΩ)	2.26	2.23	1.91	2.09	1.94	1.95	2.31	2.18	2.23	1.98
35	Parallel resistance (kΩ)	36.50	29.27	32.22	30.49	30.22	29.84	31.18	30.68	29.97	30.42
36	Leakage conductance (0.01 μ mhos)	1.62	2.00	1.91	1.82	1.92	1.90	1.78	2.02	2.01	1.87
37	Capacitance (0.01 μf)	4.36	5.53	5.01	5.21	5.16	5.20	5.29	5.39	5.55	5.12
38	Cord angle (deg)	68.88	68.31	67.65	68.08	68.81	68.95	69.12	68.74	68.61	68.50
39	Average radius (kΩ)	19.55	15.62	17.36	16.43	16.21	15.90	16.56	16.29	15.20	16.42
40	S.D. of error (kΩ)	.38	.29	.24	.26	.25	.25	.28	.24	.35	.38
41	Circle center (R) (kΩ)	20.51	16.87	18.02	17.34	17.05	16.87	17.91	17.52	16.71	17.19
42	Circle center (X) (kΩ)	6.45	4.98	6.04	5.68	6.38	5.15	5.17	5.10	5.15	5.68
	Skin impedance, resistive (kΩ)										
43	10 Hz	35.54	28.98	29.70	29.98	29.47	29.30	29.59	30.41	28.88	29.55
44	20 Hz	32.55	27.06	27.48	27.95	27.65	27.51	27.06	28.28	26.80	27.40
45	40 Hz	27.61	23.71	23.80	24.29	24.07	24.00	23.28	24.31	23.25	23.85
46	80 Hz	21.43	18.75	18.21	19.06	18.88	18.34	17.75	19.15	18.55	18.80
47	120 Hz	17.45	15.43	14.56	15.50	15.43	15.51	14.46	15.67	15.24	15.33
48	170 Hz	14.37	12.76	11.85	12.66	12.67	12.65	11.79	12.75	12.58	12.42
49	200 Hz	12.92	11.58	10.67	11.17	11.41	11.43	10.57	11.51	11.40	11.25
50	400 Hz	8.06	7.40	6.57	7.33	6.99	7.09	6.55	6.90	7.28	6.95
51	800 Hz	5.06	4.65	4.08	4.42	5.10	4.29	4.15	4.52	4.63	4.33
	Skin impedance, reactive (kΩ)										
52	10 Hz	5.45	-4.25	-4.74	-4.35	-4.20	-4.25	-4.53	-4.44	-4.16	-4.60
53	20 Hz	-8.4	-6.40	-7.09	-6.68	-6.52	-6.41	-6.96	-6.79	-6.26	-6.52
54	40 Hz	-10.93	-8.71	-9.75	-8.96	-8.89	-8.75	-9.02	-8.40	-8.40	-8.88
55	80 Hz	-12.12	-10.09	-10.93	-10.48	-10.24	-10.17	-10.78	-10.66	-9.94	-10.25
56	120 Hz	-11.83	-10.12	-10.55	-10.25	-10.29	-10.13	-10.82	-10.56	-9.83	-10.32
57	170 Hz	-11.13	-9.52	-9.98	-9.73	-9.89	-9.73	-10.22	-9.93	-9.42	-9.67
58	200 Hz	-10.68	-9.18	-9.61	-9.21	-9.43	-9.45	-9.84	-9.57	-9.09	-9.22
59	400 Hz	-7.86	-6.94	-6.96	-6.89	-6.96	-6.96	-7.24	-6.51	-6.76	-6.82
60	800 Hz	-4.45	-3.98	-3.91	-3.96	-3.86	-3.96	-4.21	-4.04	-3.39	-3.79

APPENDIX D

TABLE D-5. - VISUALLY-EVOKED RESPONSE FEATURE AVERAGES BY TASK (UNNORMALIZED DATA, N = 27)

Feature number	Feature description	Task number 1	2	3	4	5	6	7	8	9
61	RMS power	1.12	1.95	1.03	1.09	1.12	1.09	1.19	1.13	1.08
62	Overall max., amplitude*	2.74	---	2.39	2.69	2.87	2.54	2.94	2.42	2.72
63	Overal max., latency**	217	---	219	220	208	230	222	209	203
64	Overal min., amplitude	-2.58	-2.26	-2.44	-2.46	-2.49	-2.59	-2.66	-2.79	-2.38
65	Overall min., latency	197	190	181	210	216	212	191	207	215
66	Min. 100 to 160, amplitude	-2.00	-1.85	-2.03	-2.12	-2.17	-1.95	-2.18	-2.17	-1.65
67	Min. 100 to 160, latency	156	145	150	156	154	150	152	153	149
68	Max. 150 to 220, amplitude	2.35	2.36	2.05	2.17	2.22	2.10	2.43	1.91	2.47
69	Max. 150 to 220, latency	236	228	236	230	228	229	221	229	225
70	Min. 180 to 290, amplitude	-1.43	-1.19	-1.37	-1.34	-1.49	-1.62	-1.67	-1.99	-1.45
71	Min. 180 to 290, latency	331	303	296	300	295	297	303	300	299
72	Max. 215 to 270, amplitude	1.02	1.17	1.01	1.16	.95	.76	.68	.55	.94
73	Max. 215 to 270, latency	290	298	297	302	294	396	289	292	284
74	Sequential min. 1, amplitude	-1.64	-1.39	-1.76	-1.51	-1.72	-1.59	-1.66	-1.70	-1.64
75	Sequential min. 1, latency	142	131	139	130	126	139	129	135	128
76	Sequential max. 1, amplitude	.274	.362	.357	.529	.534	.335	.338	.278	.589
77	Sequential max. 1, latency	170	158	170	158	154	165	152	161	158
78	Sequential min. 2, amplitude	-1.34	-1.26	-1.12	-1.56	-1.64	-1.35	-1.53	-1.68	-.92
79	Sequential min. 2, latency	196	186	197	189	182	188	177	188	184
80	Sequential max. 3, amplitude	2.49	2.36	2.17	2.31	2.61	2.25	2.72	2.12	2.51
81	Sequential max. 3, latency	237	230	236	232	228	230	221	230	227
82	Sequential min. 3, amplitude	-1.87	-.116	-.203	-.264	-.229	-.029	-.036	-.359	-.028
83	Sequential min. 3, latency	281	272	283	275	275	268	270	274	271
84	Sequential max. 4, amplitude	.782	1.14	.816	.736	.728	.708	.694	.550	.839
85	Sequential max. 4, latency	308	297	314	303	301	291	291	296	296
86	Sequential min. 4, amplitude	-.451	-.709	-.295	-.950	-.736	-.700	-.201	-1.41	-.321
87	Sequential min. 4, latency	338	336	348	341	335	327	331	336	325
88	Number of waves	8.48	8.17	8.04	7.83	8.00	8.65	8.09	8.29	8.73

*Amplitudes are in microvolts.

**Latencies are in milliseconds.

APPENDIX D

TABLE D-6. - CORRELATION MATRIX, FINAL 10 FEATURES AND 4 CRITERIA

No.	Mean	S. D.	Feature	Description
1	. 1775	6. 9561	12	Mean interval, high
2	- . 5279	25. 5388	27	T-wave amplitude S. D. (mV)
3	- . 7093	18. 6294	31	R-T interval S. D. (seconds)
4	- . 1941	22. 6228	6	Mean amplitude, low
5	- . 3227	19. 7283	80	Sequential max. 3 (μV)
6	- . 1450	16. 4593	63	Latency overall max. (ms)
7	- . 0422	12. 7789	26	T-wave amplitude, mean (mV)
8	- . 4820	33. 0076	21	S. D. rectification pieces, high
9	. 0720	19. 3769	33	R-R interval S. D. (seconds)
10	- . 0905	17. 8781	20	Rectification, high
11	. 0011	. 7839	1	Tracking error
12	- . 0041	. 4244	2	Miss rate (percent)
13	- . 0007	. 0572	3	Response time (ms)
14	- . 0007	. 2935	4	Subjective rating (out of 60)

	12	37	31	6	80	63	26	21	33	20	1	2	3	4
12	1. 000	- . 227	- . 102	- . 357	- . 021	. 202	- . 242	- . 702	- . 152	- . 458	- . 577	- . 482	- . 442	- . 634
37	- . 227	1. 000	. 552	. 439	. 128	. 016	. 483	. 374	. 500	. 461	. 262	. 200	. 198	. 130
31	- . 102	. 552	1. 000	. 084	. 018	. 079	. 146	. 174	. 617	. 252	. 238	. 167	. 228	. 229
6	- . 357	. 439	. 084	1. 000	. 008	- . 052	. 533	. 691	. 121	. 791	. 454	. 453	. 312	. 379
80	0 . 021	. 128	. 018	. 008	1. 000	. 331	. 069	. 017	- . 102	. 088	. 185	. 122	. 150	. 131
63	. 202	. 016	. 079	- . 052	. 331	1. 000	- . 161	- . 168	. 099	- . 128	- . 196	- . 166	- . 213	- . 216
26	- . 242	. 483	. 146	. 533	. 069	- . 161	1. 000	. 386	. 007	. 467	. 346	. 189	. 169	. 282
21	- . 702	. 374	. 174	. 691	. 017	- . 168	. 386	1. 000	. 275	. 721	. 587	. 500	. 305	. 529
33	- . 152	. 500	. 617	. 121	- . 102	. 099	. 007	. 275	1. 000	. 132	. 228	. 181	. 166	. 066
30	- . 458	. 461	. 252	. 791	. 088	- . 128	. 467	. 721	. 132	1. 000	. 505	. 464	. 262	. 434
1	- . 577	. 262	. 238	. 454	. 185	- . 196	. 346	. 587	. 228	. 505	1. 000	. 712	. 686	. 802
2	- . 482	. 200	. 167	. 453	. 122	- . 166	. 189	. 500	. 181	. 464	. 712	1. 000	. 760	. 698
3	- . 442	. 198	. 228	. 312	. 150	- . 213	. 169	. 305	. 166	. 262	. 686	. 760	1. 000	. 717
4	- . 634	. 130	. 229	. 379	. 131	- . 216	. 282	. 529	. 066	. 434	. 802	. 698	. 717	1. 000

APPENDIX E
LEAST-SQUARES PREDICTION

TABLE E-1. - CUMULATIVE DISTRIBUTION OF CHI SQUARE*

Degrees of freedom	Probability of a greater value												
	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005
1	--	--	--	--	.02	.10	.45	1.32	2.71	6.84	5.02	6.63	7.88
2	.01	.02	.05	.10	.21	.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	.07	.11	.22	.35	.58	1.21	2.37	4.11	6.25	7.81	9.35	11.34	12.84
4	.21	.30	.48	.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	.41	.55	.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	.68	.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	20.84	25.34	30.43	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	21.75	26.34	31.53	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	22.66	27.34	32.62	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	23.57	28.34	33.71	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	33.66	39.34	45.62	51.80	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	42.94	49.33	56.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	52.29	59.33	66.98	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	61.70	69.33	77.58	85.53	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	64.28	71.14	79.33	88.13	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	80.62	89.33	98.64	107.56	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	90.13	99.33	109.14	118.50	124.34	129.56	135.81	140.17

*Condensed from table with 6 significant figures by Catherine M. Thompson, by permission of the Editor of Biometrika.

TABLE E-2. - CORRELATION COEFFICIENT TEST* (ENTER TABLE WITH N - 1 DEGREES OF FREEDOM)

Degrees of freedom	Level of significance		Degrees of freedom	Level of significance	
	5%	1%		5%	1%
1	.997	1.000	24	.388	.496
2	.950	.990	25	.381	.487
3	.878	..959	26	.374	.478
4	.811	.917	27	.367	.470
5	.754	.874	28	.361	.463
6	.707	.834	29	.355	.456
7	.666	.798	30	.349	.449
8	.632	.765	35	.325	.418
9	.602	.735	40	.304	.393
10	.576	.708	45	.288	.372
11	.553	.684	50	.273	.354
12	.532	.661	60	.250	.325
13	.514	.641	70	.232	.302
14	.497	.623	80	.217	.283
15	.482	.606	90	.205	.267
16	.468	.590	100	.195	.254
17	.456	.575	125	.174	.228
18	.444	.561	150	.159	.208
19	.433	.549	200	.138	.181
20	.423	.537	300	.113	.148
21	.413	.526	400	.098	.128
22	.404	.515	500	.088	.115
23	.396	.505	1000	.062	.081

*Portions of this table were taken from Table VA in Statistical Methods for Research Workers by permission of Professor R. A. Fisher and his publishers, Oliver and Boyd.

DERIVATION OF LEAST-SQUARES LINEAR PREDICTOR

If we are given N simultaneous observations on n features $(x_1, \ldots x_n)$ and a measured value of y, for those N times, we might wish to find the best linear combination of the features to predict y.

An eminently reasonable criterion to use as "best" is that the predicted value

$$\hat{y} = a_1 x_1 + \ldots + a_n x_n$$

be closest to y in a least-squares sense, i.e., that

$$\epsilon = \frac{1}{N} \sum_{i=1}^{N} (y - \hat{y})^2 = \overline{(y - \hat{y})^2}$$

be minimized.

If we write

$$a^T = [a_1 \ldots a_n]$$

then

$$\epsilon = \overline{(y - a^T x)^2} = \overline{y^2} - \overline{2ya^T x} - \overline{(a^T x)^2}$$

and it is clearly necessary that the gradient of ϵ with respect to a be zero ($\nabla_a \epsilon = 0$):

$$\nabla_a \epsilon = 0 - 2\,\overline{xy} - 2a\,\overline{xx^T}$$

Solving for a we find

$$a = \overline{xx^T}^{-1}\,\overline{xy}$$

If we consider the more general situation where there are m simultaneous measurements $(y_1, \ldots y_m)$ and we seek the best linear combination of both features and measurements

$$\hat{y} = a_1 x_1 + \ldots + a_n x_n$$

$$\tilde{y} = b_1 y_1 + \ldots + b_m y_m$$

then

$$\epsilon = \overline{(\tilde{y} - \hat{y})^2} = \overline{(b^T y - a^T x)^2}$$

There are thus m plus n necessary conditions:

$$\nabla_a \epsilon = 0 \Rightarrow a = \overline{xx^T}^{-1}\,\overline{xy^T}\,b$$

and

$$\nabla_b \epsilon = 0 \Rightarrow b = \overline{yy^T}^{-1}\,\overline{yx^T}\,a$$

Substituting the first equation into the second we have

$$b = M\,b$$

where

$$M = \overline{yy^T}^{-1}\,\overline{yx^T}\,\overline{xx^T}^{-1}\,\overline{xy^T}$$

APPENDIX E

This is an eigenvalue problem and will, under fairly general conditions, have m linearly independent solutions, i.e., there will be m scalars, λ, and m vectors b such that

$$M\,b = \lambda\,b$$

The b corresponding to the largest λ is the desired solution.

We may thus summarize the steps in solving this simultaneous least-squares prediction problem:

1) Find the mean and variance of each feature and criteria variable and standardize:

$$x_i = \frac{x_i - \bar{x}_i}{\sigma_{x_i}}, \quad y_i = \frac{y_i - \bar{y}_i}{\sigma_{y_i}}$$

2) Compute the correlation matricies:

$$xx^T,\ yy^T,\ xy^T = (y\ x\ T)^T$$

3) Invert the $n \times n$ matrix $\overline{xx^T}$ and $m \times m$ matrix $\overline{yy^T}$

4) Find the largest eigenvalue and corresponding eigenvector, b, for

$$M = \overline{yy^T}^{\,-1}\ \overline{yx^T}\ \overline{xx^T}\ \overline{xy^T}$$

5) Use this eigenvector to find the a_i's:

$$a = \overline{xx^T}^{\,-1}\ \overline{xy^T}\ b$$

REFERENCES

Baddeley, A.D., Medical Research Council Report No. A.P.U. 442, 1962. Cited by I.D. Brown, The measurements of perceptual load and reserve capacity. The Transactions of the Association of Industrial Medical Officers, 1964, 14, 44-49.

Bahrick, H.P., Nobel, M., and Fitts, P.H., Extra task performance as a measure of learning a primary task. Journal of Experimental Psychology, 1954, 48 (4), 298-302.

Benson, A.J., Huddleston, J.H.F., and Rolfe, A., Psychophysiological study at compensatory tracking on a ditial display. Human Factors, October 1965, 457-472.

Brown, I.D., Measuring the "spare mental capacity" of car drivers by a subsidiary auditory task. Ergonomics, 1962, 5, 247-250.

Brown, I.D., The measurement of perceptual load and reserve capacity. The Transactions of the Association of Industrial Medical Officers, 1964, 14, 44-49.

Brown, I.D., and Pollution, E.C., Measuring the "spare mental capacity" of car drivers by a subsidiary task. Ergonomics, 1961, 4(1), 35-40.

Cooley, W.W., and Lohnes, P.R., Multivariate Procedures for the Behavioral Sciences, p 31, Wiley, New York, 1965.

Ekstrom, P.J., Analysis of pilot workloads in flight control systems with different degrees of automation. Int. Cong. of Human Factors in Electronics, 1962.

Fraser, M.T., Aspects of the human response to high-speed, low-level flight. Aerospace Med., 33, 365-70, 1964.

Garvey, W.D. and Taylor, F.V., Interactions among operator variables, system dynamics, and task-induced stress. J. Appl. Psychol., 1959, 43, 79-85.

Glucksberg, S., Rotary pursuit tracking with divided attention to cutaneous, visual, and auditory signals. J. Engr. Psychol., 1963, 2, 119-125.

Guedry, E.F., Kennedy, S.R., Harris, S.C., and Graybiel, A.C., Human performance during two weeks in a room rotating at three rpm. Aerospace Med. 1964, 35, 1071-82.

Katz, S. and Spragg, S.D.S., Tracking performance as a function of frequency of course illumination. J. of Psychol., 1955, 40, 181-191.

Kelley, C. R., and Prosin, D. J., Adaptive techniques in measuring complete performance. Final Contract Report Nonr-4986(00), 1969.

Khalafalla, A., Turner, L., and Spyker, D., An electrical model to simulate skin dielectric dispersion. Fifth Annual Mtg., Assoc. for Advancement of Med. Instrumentation, March 1970.

Knowles, W. B., Operator Loading Tasks. Human Factors, 1963, 5, 155-161.

Knowles, W. B., and Rose, D. J., Manned lunar landing simulation. IRE Winter Convention on Military Electronics, 1963.

Kreifeldt, John G., An analysis of surface-detected EMG as an amplitude-modulated noise. Eighth ICMBE, 7.6, 1969.

Lacey, J. I., Kagan, J., Lacey, B. C., and Moss, H. A., The visceral level: situational determinant and behavioral correlates of automatic response patterns. In Expressions of the Emotions in Man (P. H. Knapp, editor), International University Press, New York, 1963, 161-208.

Levy, E. Z., Thaler, V. and Ruff, G., New techniques for recording skin resistance changes. Science 128 (1958), 33.

Lim, Thomas P. K., Cardiopulmonary function tests in clinical medicine. Thomas, Springfield, Ill., 1966.

Lindsley, D. B., Emotion. In Handbook of Experimental Psychology (S. S. Stevens, editor) John Wiley, New York, 1951.

Martin. R. J., and Levy, M. N., Effects of sympathetic and parasympathetic interaction on ventricular function. Int. Conf. on Med and Biol. Eng., p. 285, 1967.

McDonnell, J. D., An application of measurement methods to improve the quantitative nature of pilot rating scales. IEEE Transactions on Man-Machine Systems, Vol. MMS-10, No. 3, September, 1969.

McDonnell, J. D., Pilot rating techniques for the estimation and evaluation of handling qualities. Technical Report AFRDC-TR-68-76, December 1968.

Pribram, K. H., Emotion: steps toward a neuropsychological theory. In Neurophysiology and Emotion. The Rockefeller University Press and Russell Sage Foundation, New York, 1967, 3-40.

Schaefer, H. and Haas, H. G., Electrocardiography. In Handbook of Physiology, Circulation I, 342, American Physiological Society, 1965.

Senders, J.W., Tracking with intermittently illuminated displays. Wright Air Development Center, Technical Report No. WADC TR 55-378, October 1955.

Shaw, W.J., The human factor in motor traffic. International Road Safety and Traffic Review, 1957, 10, 37-39. Cited by I.D. Brown, Measuring the "spare mental capacity" of car drivers by a subsidiary auditory task. Ergonomics, 1962, 5(1), 247-250.

Sinbel, A., In vivo and in vitro measurement of the impedance and phase characteristics of human tissues by mutual impedance methods. PhD Thesis, University of Minnesota, 1966.

Smith, M.H., Jr., Spread of effect is the spurious result of nonrandom response tendencies. Journal of Experimental Psychology, 1949, 39, 355-369.

Snedecor, G.W., and Cochran, W.G. Statistical Methods, Iowa State University Press, p 134, 1967.

Spyker, D.A., Cardiovascular control and the artificial heart. Biomedical Sciences Instrumentation, 7, Instrument Society of America, 110, 1970.

Soliday, M.S. and Schohan, B., Performance and physiological responses of pilots in simulated low-altitude, high-speed flight. Aerospace Med. 1965, 36, 100-04.

Tune. G.S., A brief survey of variables that influence random generation. Perceptual and Motor Skills, 1964, 18, 705-710.

Warner, H.R., Control of heart function", Ann. Conf. on Eng. in Med. and Biol. p. 116, 1967.

Wee, W.G., Generalized inverse approach to adaptive multiclass pattern classification. IEEE Trans. on Computers, 1968, Vol. C-17, No. 12, p 1157.

Wee, W.G., On feature selection in a class of distribution-free pattern classifiers, IEEE Trans. on Information Theory, January 1970.

Zeitlin. L.R., Development of a "spare capacity" measure of operator loading on sensor motor tasks. Paper presented at the meeting of the American Psychological Association, New York, September 1966.

NASA-Langley, 1971 —— 5